The Adenoviruses

Edited by
HAROLD S. GINSBERG
College of Physicians and Surgeons of Columbia University
New York, New York

PLENUM PRESS • NEW YORK AND LONDON

Library of Congress Cataloging in Publication Data

Main entry under title:

The Adenoviruses.

(The Viruses)
Includes bibliographical references and index.
1. Adenoviruses. I. Ginsberg, Harold S., 1917– . II. Series.
QP396.A34 1984 576'.64 84-8264
ISBN 0-306-41592-5

CHAPTER 3

The Structure of the Genome

JOHN S. SUSSENBACH

I. INTRODUCTION

Adenovirus particles have a highly ordered structure and are composed of protein and DNA. Human adenoviruses contain about 87% protein and 13% DNA (Green and Piña, 1963), while the larger avian chick embryo lethal orphan (CELO) virus consists of 83% protein and 17% DNA (Laver *et al.*, 1971). In virions, the viral DNA is tightly associated with several virus-coded proteins. Disruption of virions with acetone, urea, or pyridine, or repeated freezing and thawing, releases the viral cores, which, in addition to the viral DNA, still contain about 18–20% of the total protein of the virions (Laver *et al.*, 1967, 1968; Maizel *et al.*, 1968; Prage *et al.*, 1968, 1970). The proteins found in viral cores are mainly two basic polypeptides. The major core protein is identical to polypeptide VII [molecular weight 18,000 (18K)], of which about 1000 copies are present in each viral particle. The minor core protein is polypeptide V (molecular weight 45.5K), of which each virion contains about 200 copies (Laver *et al.*, 1968; Prage *et al.*, 1968, 1970; Prage and Pettersson, 1971; Russell *et al.*, 1971; Everitt *et al.*, 1973; Laver, 1970). However, when cores are prepared by extraction of virions with sarkosyl, only polypeptide VII is found associated with the DNA (Brown *et al.*, 1975). The different protein compositions of pyridine and sarkosyl cores suggest that polypeptide VII is more intimately associated with the viral genome than is polypeptide V.

Corden *et al.* (1976) concluded that adenovirus DNA packed in virions has a chromatinlike structure. They found that digestion of disrupted virions with micrococcal nuclease cleaves the viral genome into fragments about 200 nucleotides long. However, these experiments could

JOHN S. SUSSENBACH • Laboratory for Physiological Chemistry, State University of Utrecht, 3521 GG Utrecht, The Netherlands.

not be repeated by Tate and Philipson (1979). Mirza and Weber (1982) proposed that although adenovirus DNA is indeed packed into subunits, its organization in the virion is not completely the same as that of eukaryotic chromatin. Partial deoxyribonuclease (DNase) digestion of eukaryotic chromatin leads to stretches of DNA with a length of 200 nucleotide pairs associated with histones. Mirza and Weber (1982) found that viral chromatin does indeed have a nucleosomelike structure, but that partial DNase digestion yields monomers of about 150 nucleotide pairs of DNA wrapped around three dimers of polypeptide VII. These monomers are linked by a variable length of DNA associated with one copy of polypeptide V.

Since adenovirus DNA is tightly associated with virion proteins, protein-free DNA can be obtained only by extensive digestion of virions or viral cores with proteolytic enzymes (papain, pronase, or proteinase K) followed by sodium dodecyl sulfate (SDS)–phenol extraction (van der Eb and van Kesteren, 1966; Green *et al.*, 1967; van der Eb *et al.*, 1969; Laver *et al.*, 1971). The DNA thus isolated has a linear structure and has been characterized in great detail.

An alternative isolation procedure for adenovirus DNA was first applied by Bellett and co-workers for CELO and adenovirus type 2 (Ad2) DNA (Robinson *et al.*, 1973; Robinson and Bellett, 1975a). These investigators isolated DNA in the absence of proteolytic enzymes, employing an extraction with 4 M guanidinium hydrochloride. The isolated DNA has in the electron microscope (EM) a circular structure, which can be converted into a linear configuration by digestion of the preparation with proteolytic enzymes (Robinson *et al.*, 1973). Similar studies have also been performed for Ad5 DNA (Keegstra *et al.*, 1977). The sensitivity of the circular structures for proteolytic enzymes suggests that the circular structures are maintained by a protein linker.

By *in vitro* labeling of the protein moiety with ^{125}I, it could be demonstrated that a polypeptide with a molecular weight of 55K is covalently attached to the 5′ end of each DNA strand (Rekosh *et al.*, 1977). This protein, designated terminal protein, has a hydrophobic character, which facilitates joining of the ends of the DNA–protein complexes, resulting in the formation of circular structures and concatemers. The properties of the linear deproteinized DNA as well as the characteristics of the circular DNA–protein complexes are discussed in more detail in the following sections.

II. GROUPING OF ADENOVIRUSES BASED ON DNA HOMOLOGY

The different human adenoviruses have been classified into subgroups on the basis of different criteria. Rosén (1960) originally proposed three subgroups based on differences in hemagglutinating capacity.

Hierholzer (1973) extended this classification system to ten subgroups. On the basis of the apparent molecular weights of virion polypeptides V, VI, and VII, Wadell (1978) arranged 20 human serotypes into five groups. A completely different type of classification is based on the oncogenicity of the human adenoviruses. The different serotypes have been subdivided into a highly oncogenic subgroup A (Ad12, Ad18, Ad31), a weakly oncogenic subgroup B (e.g., Ad3 and Ad7), and a nononcogenic subgroup C (e.g., Ad2 and Ad5) (Trentin *et al.*, 1962; Girardi *et al.*, 1964; Huebner *et al.*, 1962, 1965; Larson *et al.*, 1965; Pereira *et al.*, 1965; Green, 1970). It is interesting to note that there is a correlation between the guanine–cytosine (GC) content of the human adenovirus DNAs and the oncogenicity of the viruses. The GC content of the DNAs decreases with increasing oncogenicity (Piña and Green, 1965) (Table I). Probably this correlation has no physiological basis, since, in contrast to the human adenoviruses, the oncogenic simian adenoviruses tend to have slightly higher GC contents than the nononcogenic adenoviruses (Goodhearst, 1971). Further, the oncogenic simian serotypes have GC contents that are in general higher than those of the nononcogenic human serotypes.

The most meaningful and fundamental way to group adenoviruses is based on DNA sequence homology. Fortunately, the DNA homology grouping is in agreement with other groupings of human adenoviruses on the basis of oncogenicity, GC content, and molecular characteristics of viral proteins (Table I). Originally, Green *et al.* (1970) determined the homology among different DNAs employing filter hybridization. Recently, the classification was improved by employment of liquid-phase molecular hybridization with *in vitro*-labeled viral DNA. A total of 31 different human adenovirus serotypes were divided into five different subgroups, A–E (Green *et al.*, 1979b). In general, members of the same subgroup have genomes that are homologous for more than 90%. However, members of subgroup A share only 48–69% of their DNA sequences. The homology among members of different subgroups is less than 20% (Table I).

The major regions of least homology among DNAs of different human serotypes have been visualized by heteroduplex mapping (Garon *et al.*, 1973). Heteroduplexes of subgroups B and C DNAs contain two major regions of heterology located at positions 50–65 and 78–91 on the adenovirus genome map. Heteroduplexes of members of subgroup A show a more complex distribution of homologous and heterologous regions. However, in this case, too, heterology is found at the two positions mentioned above.

Using the single-strand specific endonuclease from *Neurospora crassa*, Bartok *et al.* (1974) were able to digest specifically the heterologous regions from heteroduplexes of Ad2 and Ad5 DNA and obtained three specific fragments, in agreement with the heteroduplex mapping. The heterologous regions contain the genetic information of the major coat proteins hexon and fiber, which play an important role in the se-

TABLE I. Properties of Human Adenovirus DNA Homology Groups A–E[a]

Groups	Types	DNA homology[b]	DNA MWs ($\times 10^{-6}$)[c]	DNA GC (%)[d]	ITR length[e]	Tumor induction[f]	Cell transformation[g]	HA group[h]
A	12, 18, 31	48–69% within group, 8–20% with other types	19.2–22.0	47–49	162/164 for Ad12, 165 for Ad18	High	+	3B
B	3, 7, 11, 14, 16, 21	89–94% within group, 9–20% with other types	22.7–23.0	49–52	136 for Ad3 and Ad7	Weak	+	1A, B
C	1, 2, 5, 6	99–100% within group, 10–16% with other types	23.0	57–59	102/103 for Ad2, 103 for Ad5	Nil	+	3A
D	8–10, 13, 15, 17, 19, 20, 22–30, 32, 33, 36, 37	94–99% within group, 4–17% with other types	?	57–59	?	Nil	?	2A–F
E	4	4–23% with other types	22.8	High	116	Nil	?	3A

[a] (MW) Molecular weight; (ITR) inverted terminal repetition; (HA) hemagglutination.
[b] Data from Green *et al.* (1979b).
[c] Data from Green and Piña (1964), van der Eb and van Kesteren (1966), and Tibbetts (1977).
[d] Data from Piña and Green (1965).
[e] Data from Steenbergh *et al.* (1977), Arrand and Roberts (1979), Shinagawa and Padmanabhan (1979), (1980), Tokunaga *et al.* (1982), Sugisaka *et al.* (1980), Garon *et al.* (1982), and Schwarz *et al.* (1982).
[f] Data from Girardi *et al.* (1964), Huebner *et al.* (1962, 1965), Larson *et al.* (1965), Pereira *et al.* (1965), and Trentine *et al.* (1962).
[g] Data from McBride and Wetner (1964), Sekikawa *et al.* (1978), Freeman *et al.* (1967), Gallimore (1974), van der Eb *et al.* (1977), and McAllister *et al.* (1969).
[h] Data from Hierholzer (1973).

rological classification of the different adenovirus serotypes. In addition, one of the heterologous regions codes for a group of nonvirion early proteins (see Section VII).

III. PHYSICOCHEMICAL PROPERTIES OF ADENOVIRUS DNA

DNA, extracted from adenovirus particles employing digestion with proteolytic enzymes, has a linear double-stranded structure (van der Eb and van Kesteren, 1966; Green *et al.*, 1967; van der Eb *et al.*, 1969; Younghusband and Bellett, 1971). The size of the viral genome varies from serotype to serotype. The molecular weights of the human adenovirus DNAs range from 19–22 $\times$ 10^6 for the highly oncogenic serotypes Ad12, Ad18, and Ad31 to 23–24 $\times$ 10^6 for the nononcogenic serotypes Ad1, Ad2, and Ad5 (Green *et al.*, 1967) (Table I). On the basis of nucleotide sequence data and the sum of restriction fragments, it has been inferred that the genome of Ad2 and Ad5 is about 36,000 nucleotide pairs and that Ad12 DNA is 34,300 nucleotide pairs long. The sizes of the genomes of nonhuman serotypes are comparable to those of their human counterparts [that of mouse serotype FL DNA being 20.7 $\times$ 10^6 (Temple *et al.*, 1981) and of simian adenovirus SA7 DNA being 22 $\times$ 10^6 (Burnett and Harrington, 1968)]. On the other hand, the genome of the avian chick embryo lethal orphan (CELO) virus is much larger, measuring 30 $\times$ 10^6 (Younghusband and Bellett, 1971; Laver *et al.*, 1971).

When native adenovirus DNA is digested with *Escherichia coli* exonuclease III and is subsequently examined under the EM, no circularization of the linear genome is observed, indicating that adenovirus DNA is not terminally redundant as T7 DNA (Green *et al.*, 1967; Younghusband and Bellett, 1971). On the other hand, when double-stranded DNA (dsDNA) is denatured and reannealed at low DNA concentrations, both strands of human as well as of avian adenovirus DNA are able to form single-stranded circles (Garon *et al.*, 1972; Wolfson and Dressler, 1972; Robinson and Bellett, 1975b). The formation of single-stranded circles indicates that adenovirus DNA contains an inverted terminal repetition. This inverted terminal repetition is discussed in more detail in Section V.

The distribution of adenine–thymine (AT) and GC base pairs in adenovirus DNA has been investigated by partial thermal denaturation mapping. The unique thermal denaturation patterns of DNAs from Ad2, Ad5, and Ad12, the avian CELO virus, and the mouse strain FL indicate that adenovirus DNA is not circularly permuted as T7 DNA, but that all DNA molecules from the same serotype have an identical nucleotide sequence (Doerfler and Kleinschmidt, 1970; Younghusband and Bellett, 1971; Doerfler *et al.*, 1972; Ellens *et al.*, 1974; Temple *et al.*, 1981). In most denaturation patterns, the distribution of AT and GC base pairs

along the DNA molecule is asymmetrical. By convention, the AT-rich half of an adenovirus DNA molecule has been designated the right-hand half of the molecule (Doerfler and Kleinschmidt, 1970). In some cases (Ad2 and Ad5), the AT- and GC-rich halves of the DNA molecules can be separated by CsCl or $HgCl_2$–Cs_2SO_4 gradient centrifugation of sheared DNA (Kimes and Green, 1970; Doerfler and Kleinschmidt, 1970; Horwitz, 1974; Graham *et al.*, 1974b). However, due to the more even distribution of AT and GC base pairs in Ad12 DNA, separation of the left and right halves of Ad12 DNA by this procedure is not possible (Doerfler *et al.*, 1972).

Separation of the complementary strands of adenovirus DNA can be performed by complexing of the single strands of denatured native DNA with poly(I:G) or poly(U:G). Intact complementary strands have been obtained for Ad2, Ad5, Ad7, and Ad12 DNA (Kubinski and Rose, 1967; Landgraf-Leurs and Green, 1971; Patch *et al.*, 1972; Tibbetts *et al.*, 1974; Vlak *et al.*, 1975). Since the two complementary strands bind unequal amounts of the copolymers, the two strands can be separated by equilibrium density-gradient centrifugation or by gel electrophoresis (Goldbach *et al.*, 1978). Complementary strands of Ad2 and Ad5 DNA have also been separated by alkaline CsCl equilibrium density-gradient centrifugation (Sussenbach *et al.*, 1973; Sharp *et al.*, 1975). The buoyant densities of the two strands in alkaline CsCl differ by 2–4 mg/ml, which is sufficient for separation. The heavy strands of Ad2 and Ad5 DNA obtained by poly(U:G)–CsCl gradient centrifugation have the lower density in alkaline CsCl (Tibbetts *et al.*, 1974; Vlak *et al.*, 1975).

Tibbetts *et al.*, (1973) showed that Ad2 single-stranded DNA (ssDNA) is retained by hydroxyapatite columns under conditions generally used for selective retention of dsDNA, probably due to partialy complementary regions in the single strands. Other indications for regions of complementarity in adenovirus ssDNA were obtained by EM. Under suitable conditions, an extended region of secondary structure is observed at position 73 on the conventional adenovirus map (Wu *et al.*, 1977). Regions that contain complementary sequences were also detected at the molecular termini (Padmanabhan and Green, 1976; Wu *et al.*, 1977). Digestion of native Ad2 DNA with exonuclease III followed by repair synthesis of the exposed single-stranded ends with DNA polymerase I revealed the presence of self-complementary sequences about 50 nucleotides long, located at a distance of about 180 nucleotides from each molecular end (Padmanabhan and Green, 1976). Nucleotide sequence analysis of the termini confirmed the existence of self-complementary sequences in these regions.

IV. COORDINATE SYSTEM

To come to an unambiguous nomenclature for the two complementary strands of adenovirus DNA, it has been proposed to adopt a nomen-

clature that is based on the direction of transcription, rather than on physical properties, e.g., densities. By convention, the AT-rich half of the DNA molecule is oriented to the right and the strand transcribed to the right is called the r-strand, while the leftward-transcribed strand is designated the l-strand.* The r-strand appears to be identical to the strand with the higher density in alkaline CsCl and to the strand with lower density in poly(U:G)–CsCl (see the proposal in *J. Virol.* 22:830, 1977). Further, it is agreed to divide the adenovirus DNA into 100 map units (m.u.) from left to right on the viral genome.

The agreement on a unique orientation of adenovirus DNA molecules formed the basis for an unambiguous mapping of significant landmarks on the adenovirus genome. With the discovery and the purification of restriction endonucleases, powerful tools became available to dissect the adenovirus genome in distinct specific fragments (for a review of available enzymes, see Roberts, 1981). These fragments have been used to unravel the organization of the adenovirus genome in detail. For many adenovirus serotypes, accurate restriction endonuclease cleavage maps of the viral genome are available, and with the increasing knowledge of the nucleotide sequences of several adenovirus DNAs, this number is still growing. A summary of restriction endonuclease cleavage maps is presented in Appendix A.

Many restriction fragments have been inserted into prokaryotic plasmids employing recombinant DNA techniques (Stenlund *et al.*, 1980). These adenovirus DNA-containing plasmids are very useful for obtaining large amounts of specific fragments, especially of poorly growing serotypes. They have frequently been used for nucleotide sequence analysis and site-directed mutagenesis. The two complementary strands of restriction fragments have been separated by annealing denatured fragments in the presence of an excess of one of the intact complementary strands followed by separation of the partial duplex and the remaining single strand. Strand separation has also been obtained by gel electrophoresis of denatured restriction fragments (Tibbetts and Pettersson, 1974; Sharp *et al.*, 1975; Sussenbach *et al.*, 1973; Goldbach *et al.*, 1978). These single strands have frequently been used to isolate specific messenger RNA (mRNA) species.

The most detailed information on the structure of the adenovirus genome and the positions of important landmarks became available by nucleotide sequence analysis of DNAs from different adenovirus serotypes (see Appendix B). The most extended sequences have been established for Ad2 DNA, of which about 70% has been sequenced (Arrand and Roberts, 1979; Zain and Roberts, 1979; Zain *et al.*, 1979a,b; Shinagawa and Padmanabhan, 1979; Galibert *et al.*, 1979; Akusjärvi and Pettersson, 1978a,b, 1979a,b; Hérissé *et al.*, 1980, 1981; Akusjärvi *et al.*,

* It should be noted that r-strand transcripts are equivalent to l-strand DNA sequences and that l-strand transcripts are homologous to r-strand sequences.

1980, 1981; Shinagawa *et al.*, 1980; Hérissé and Galibert, 1981; Alestrom *et al.*, 1980, 1982; Akusjärvi and Persson, 1981a; Kruijer *et al.*, 1982; Gingeras *et al.*, 1982). This allows the positioning of many landmarks on the Ad2 genome at the nucleotide level. Comparison of the Ad2 nucleotide sequence and the restriction maps revealed that the nucleotide equivalent of 1% of the genome depends on the particular location on the Ad2 genome (Gingeras *et al.*, 1982). It was derived that a value of 365 nucleotides for 1% gives the best fit for the left end, while a value of 357 nucleotides for 1% is the best fit for the right end. The differences in nucleotide equivalent for 1% are probably caused by the differences in nucleotide composition between the right and left halves of the Ad2 genome.

V. INVERTED TERMINAL REPETITION

The existence of an inverted terminal repetition (ITR) in adenovirus DNA was discovered when denatured DNA was reannealed at low concentrations and examined under the EM. A high percentage of the single strands were present in a circular form, indicating that adenoviral DNA contains an ITR (Garon *et al.*, 1972; Wolfson and Dressler, 1972). So far, ITRs have been detected in every serotype investigated, although the length of the repetitions may vary (Table I). The general occurrence of an ITR in adenovirus DNA suggests very strongly that this feature plays an important role in viral propagation.

The single-stranded circular structures have a rather high thermal stability, which is consistent with a highly ordered base-pairing between the terminal sequences (Garon *et al.*, 1972; Wolfson and Dressler, 1972). It also suggests that the ITRs must be of considerable length. Circularization of adenovirus ssDNA can be abolished by digestion with exonuclease III, and this treatment has been used to estimate the size of the terminal repetitions. Garon *et al.* (1972) concluded that the length of the terminal repetition ranged from 350 base pairs (bp) for Ad2 to 1400 bp for Ad31. However, since inverted repeats of these sizes can be visualized under the EM and no double-stranded regions were detected in the single-stranded circles, it was concluded that the exonuclease III experiments obviously lead to an overestimation of the lengths of the ITRs. An exceptionally long ITR was detected in Ad18 DNA (Garon *et al.*, 1975). In single-stranded circles of this serotype, a double-stranded panhandle with a mean length of 0.31 μm was seen, equivalent to 3% of the genome length.

A more accurate estimate of the size of the ITR of Ad2 DNA was obtained by restriction enzyme analysis of end-labeled DNA. When a restriction enzyme cleaves within the repeated sequence, both molecular ends will yield a fragment of the same size, while cleavage outside the repeated sequence will yield fragments of different size. Employing this

approach, Roberts *et al.* (1974) estimated that the terminal repetition of Ad2 DNA is between 100 and 140 nucleotides long (also see Arrand *et al.*, 1975).

Recently, nucleotide sequence analysis has been used to determine exactly the size and composition of several adenovirus serotypes (Appendix B). Some general features of the adenovirus ITRs can be demonstrated in the ITR of Ad5 DNA, the first sequenced repetition. The ITR of Ad5 is 103 bp long (Steenbergh *et al.*, 1977). Its sequence is unique and does not contain extended self-complementary regions. A striking property of the Ad5 terminal repetition is the asymmetrical distribution of GC and AT base pairs. The first 50 bp contain 72% AT, while the next 50 bp have only 27% AT. Although the lengths of inverted repeats of other serotypes may differ considerably, they all show the same asymmetrical distribution of base pairs. As for a function of this property, it is not unlikely that the high AT content of the first half of terminal repetitions is of relevance for a rapid unwinding of the molecular ends during initiation of DNA replication.

Comparison of the inverted repetitions of serotypes from the same subgroup shows a high degree of homology (see Appendix B). The repetitions of Ad2 and Ad5 both have a length of 103 bp and are completely identical (Steenbergh *et al.*, 1977; Shinagawa and Padmanabhan, 1979), although the repetition of a particular Ad2 strain has been described that is 102 bp long (Arrand and Roberts, 1979). The terminal repetitions of Ad3 and Ad7 strain Greider both have a length of 136 bp and differ at 7 positions (Tolun *et al.*, 1979; Shinagawa and Padmanabhan, 1980). Comparison of two Ad7 strains (Greider and Gomen) reveals that both repeats are 136 bp long but differ at 5 positions (Dijkema and Dekker, 1979; Shinagawa and Padmanabhan, 1980). Similar strain differences have also been found for Ad12. The length of the Ad12 ITR varies between 162 (Shinagawa and Padmanabhan, 1980) and 164 bp (Sugisaki *et al.*, 1980; Schwarz *et al.*, 1982). In all ITRs determined except one, a dCMP residue has been found at the 5′ ends of adenovirus DNA. The exception is chick enbryo lethal orphan (CELO) DNA, which has at its 5′ end a dGMP residue (Alestrom *et al.*, 1982a). In the ITRs of all human adenovirus DNAs, the sequence ATAATATACCTTAT (nucleotides 9–22) is present (Tolun *et al.*, 1979); the regions of the inverted repetitions beyond nucleotide 50 show a low degree of homology, although in all serotypes an asymmetrical distribution of base pairs is found. Comparison of the DNAs of the human serotypes with mouse strain FL DNA (Temple *et al.*, 1981) reveals that they have the sequence ATAATATAC (nucleotides 9–17) in common, while the homologous region between human adenovirus DNAs and CELO DNA is located between positions 9 and 15 (ATAATAT) (Alestrom, *et al.*, 1982a). It is very likely that the conserved sequences 9–15 and 9–17 play a crucial role in the initiation of DNA replication and are probably involved in recognition of the site of initiation by the precursor of the terminal protein. In this respect, it is interesting to note that mouse

adenovirus strain FL DNA can be replicated in an *in vitro* DNA replication system of Ad2 DNA (Temple *et al.*, 1981). Shinagawa and Padmanabhan (1980) have pointed out that in Ad2, Ad3, Ad5, Ad7, and Ad12 DNA, an additional region of interesting homology is present. In these serotypes, the hexanucleotide TGACGT is found at or near the site where the sequences beyond the ITR begin to diverge. The function of this homology is unknown.

VI. TERMINAL PROTEIN

The presence of protein at the termini of adenovirus DNA was originally detected by Bellett and co-workers, employing DNA isolation procedures that avoid proteolytic digestion (Robinson *et al.*, 1973; Robinson and Bellett, 1975a). These investigators observed that the DNA–protein complex obtained is resistant to boiling and treatment with SDS, indicating that the protein is probably covalently linked to the DNA (Robinson *et al.*, 1973; Sharp *et al.*, 1976; Carusi, 1977; Padmanabhan and Padmanabhan, 1977).

When the buoyant densities of Ad2 and Ad5 DNA–protein complexes are compared with the densities of the corresponding DNAs isolated by digestion with pronase, a small difference of 2–10 mg/ml is found. This corresponds to an amount of protein present in the DNA–protein complex of a maximal 0.3% of the total virion protein (Robinson and Bellett, 1975a; Keegstra *et al.*, 1977). By gel electrophoresis of labeled DNA-free terminal protein (TP), it could be established that TP has an apparent molecular weight of 55K (Rekosh *et al.*, 1977).

Due to the hydrophobic character of TP, DNA–protein complexes aggregate very easily. As a result of this aggregation, DNA–protein complexes accumulate on tops of agarose and polyacrylamide gels during electrophoresis. It has been observed that when DNA–protein complexes are digested with restriction endonucleases and the digestion products are separated by gel electrophoresis, the terminal fragments carrying TP preferentially stay on top of the gel, while internal fragments conventionally run into the gel (Brown *et al.*, 1975; Sharp *et al.*, 1976). Another way to separate the DNA–protein complexes from protein-free DNA is based on differential binding of these compounds to glass-fiber filters (Coombs and Pearson, 1978; Coombs *et al.*, 1978).

To establish the nature of the DNA–protein linkage, deproteinized DNA and DNA–protein complexes have been subjected to enzymatic and nonenzymatic treatments. Both types of DNA are inaccessible to phosphatase, DNA polynucleotide kinase, and λ-exonuclease VII (Carusi, 1977; Sharp *et al.*, 1976), indicating that the 5′ ends of adenovirus DNA are blocked. On the other hand, the 3′ ends can freely be labeled with terminal transferase and are accessible to exonuclease III. These results are most easily explained assuming that in the DNA–protein complex,

TP is covalently attached to the 5′ ends of the two complementary strands. The inaccessibility of deproteinized DNA is probably due to the fact that the 5′ ends are still linked to short peptides. Treatment of DNA–protein complexes or deproteinized DNA with alkali or piperidine removes these peptides and makes the DNA freely accessible for enzymes (Robinson *et al.*, 1973; Carusi, 1977; Tolun *et al.*, 1979; Rekosh, 1981). TP can also be separated from adenovirus DNA by digestion with nuclease S1 (Ariga *et al.*, 1979; Roninson and Padmanabhan, 1980; Rijnders *et al.*, 1983). The DNA–protein complex is cleaved in close proximity to the protein–DNA linkage and yields a protein with a molecular weight of 55K (Rijnders *et al.*, 1983). Recently, Rekosh (1981) showed that treatment of the Ad2 DNA–protein complex with piperidine releases a protein with a molecular weight of 52K. This observation suggests that after DNase I or S1 digestion, the TP isolated still contains a few nucleotide residues.

The nature of the linkage between TP and the DNA molecule has been elucidated by Desiderio and Kelly (1981). Their experiments clearly indicate that Ad2 TP is bound to DNA by a phosphodiester bond between the hydroxyl group of a Ser residue of TP and the 5′-phosphate group of the terminal deoxycytidine residue of the two complementary strands of adenovirus DNA. The particular Ser residue in the TP amino acid sequence involved in the linkage of TP to DNA has recently been identified (Smart and Stillman, 1982).

The origin of TP has been uncertain for many years. Green *et al.* (1979c) showed by tryptic fingerprinting of TPs of five different human serotypes that these proteins were very similar in structure. On the other hand, Rekosh (1981) found different sizes for the TPs of different human serotypes, suggesting that TP is not of cellular origin. He concluded that TP is a highly conserved virus-coded protein. The viral origin of TP was unambiguously proved by Stillman *et al.* (1981), who showed that cell-free translation of mRNAs selected from a region between coordinates 11 and 31.5 on the viral l-strand (see Section IV) leads to synthesis of proteins with apparent molecular weights of 105, 87, and 75K. The 87K protein appeared to be identical to an 80K protein (Challberg *et al.*, 1980) that is covalently attached to the 5′ ends of growing Ad2 DNA strands synthesized in an *in vitro* DNA replication system (Challberg and Kelly, 1979a,b). The 80K protein is structurally related to TP, suggesting that TP is synthesized as an 80K precursor TP (pTP) and that pTP is the active form of TP in adenovirus DNA replication. The different molecular weights found for pTP (80 and 87K) are due to the use of different molecular-weight markers. The 80/87K protein appears to be identical to the protein that is covalently attached to the DNA from temperature-sensitive (*ts*) mutant Ad2*ts*1 virions grown at the nonpermissive temperature (Stillman *et al.*, 1981; Challberg and Kelly, 1981). Ad2*ts*1 is a mutant that cannot cleave virus-coded precursor proteins to their mature counterparts during virion maturation (Bégin and Weber, 1975; Weber *et al.*, 1975).

The mapping of pTP on the virus genome led to the definition of a new early transcription unit, designated E2b. The structure of this region is discussed in detail in Section VII.B.3.

Evidence has been presented that TP plays an essential role in the initiation of adenovirus DNA replication. Analysis of the *in vitro* DNA replication system developed by Challberg and Kelly (1979a,b), in which the DNA–TP complex is used as a template, showed that the first step in the replication of adenovirus DNA is the linkage of dCMP to pTP. The protein probably recognizes a specific sequence within the inverted terminal repetition, which might be involved in binding of pTP to the DNA (Tamanoi and Stillman, 1982). It is likely that the conserved sequence 9–22 in different adenovirus serotypes functions as such a recognition sequence. The presence of TP in the DNA–TP complex might stabilize the initiation complex. Recently, it was shown that the protein is dispensable (Tamanoi and Stillman, 1982), since adenovirus DNA devoid of TP or remaining amino acids can also be used as template in an *in vitro* DNA replication system. It has been proposed that the presence of TP in the DNA–TP complex protects the viral DNA against nucleolytic degradation.

A protecting function of TP has also been proposed to explain the high infectivity of DNA–protein complexes. Deproteinized DNA is infectious when assayed by the calcium coprecipitation procedure (Nicolson and McAllister, 1972; Graham and van der Eb, 1973). However, the infectivity of DNA–TP complexes is 50–100 times higher (Sharp *et al.*, 1976; Chinnadurai *et al.*, 1978; van Wielink, 1978). Although the difference in infectivity might be due to a protective function of TP, it cannot be excluded that the presence of TP on the template is essential for accurate positioning of the pTP on the DNA during the first stage of initiation of adenovirus DNA replication. The role of TP in DNA replication is discussed extensively in Chapter 7.

VII. ORGANIZATION OF THE ADENOVIRUS GENOME

For the unraveling of the organization of the adenovirus genome, a great variety of techniques have been employed, i.e., DNA–RNA hybridization, R-loop mapping, genetic mapping of mutants, translation of preselected mRNA species, and nucleotide sequence analysis (for details, see Mautner *et al.*, 1975; Sambrook *et al.*, 1975; Grodzicker *et al.*, 1975, 1977; Chow *et al.*, 1977b, 1979a,b; Berk and Sharp, 1977a, 1978; Westphal *et al.*, 1976; Westphal and Lai, 1977; Kitchingman *et al.*, 1977; Kitchingman and Westphal, 1980; Miller *et al.*, 1980) (for sequences, see Appendix B). Despite a substantial nucleotide sequence divergence, all adenovirus serotypes studied so far show the general genetic organization (see Appendix B). Since the genomes of the highly homologous types Ad2 and Ad5 have been investigated most extensively, the organization of the

adenovirus genome is discussed employing for the most part data obtained with these particular serotypes. The precise location of major landmarks at the nucleotide level is indicated in the Ad2 sequence (Appendix B), unless otherwise stated. During the productive infection cycle of adenoviruses, the different viral genes are expressed in a rather complex pattern (Tooze, 1981; Persson and Philipson, 1982).

Traditionally, the adenovirus genes are subdivided into early genes, which are expressed before the onset of viral DNA replication, and late genes, which are transcribed after replication of adenovirus DNA has started. However, a group of intermediate genes has also been distinguished. These genes are expressed at intermediate times in infection in the absence of DNA synthesis and are also easily detected at late times. The complex transcription pattern of adenovirus DNA is discussed extensively in Chapter 5. A summary of the major RNA transcripts and the corresponding proteins is presented in Figs. 1 and 2. These diagrams demonstrate that the adenovirus genetic information is scattered over the

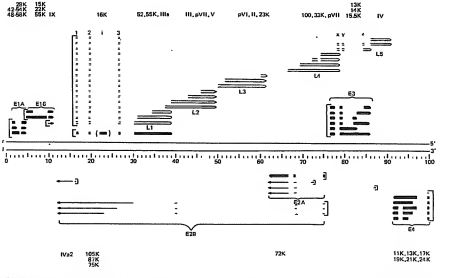

FIGURE 1. Transcriptional organization of the Ad2 genome. The genome is divided into 100 map units. The r-strand is rightward-transcribed into RNA and the l-strand leftward. The direction of transcription is indicated by arrows. The capped 5′ ends of the cytoplasmic RNA indicate the positions of transcriptional promoters, while the arrowheads represent the 3′ polyadenylation sites. Gaps in arrows indicate intervening sequences, which have been removed from the cytoplasmic RNA by splicing. The RNA shown in bold lines can be detected early in infection before the onset of DNA replication (regions E1a, E1b, E2a, E3, E4; also the late promoter at 16.5 units is active early in infection, leading to transcription to 39 units). The light lines represent intermediate RNAs synthesized at early as well as at late times in the infection cycle (E2a, E2b, polypeptide IX). The double-lined arrows indicate late RNA species. Correlations of mRNAs with encoded proteins are based on cell-free translation of selected RNA species and RNA mapping data. Proteins are designated by their molecular weights in kilodaltons (K) or by roman numerals (virion components).

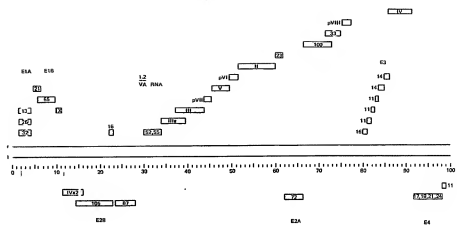

FIGURE 2. Protein-coding regions of the Ad2 genome. The regions on the adenovirus genome that code for protein have been determined by hybrid-arrest translation, by *in vitro* translation of preselected mRNAs, by RNA mapping, and by direct DNA and RNA sequence analysis. The identified proteins are designated by their apparent or theoretical molecular weights in kilodaltons or by roman numerals (virion components). Regions pVI, pVII, and pVIII indicate the positions of the precursors of polypeptides VI, VII, and VIII. Interrupted coding regions indicate discontinuous genes.

two complementary strands. About 69% of all genetic information is located on the rightward-transcribed strand (r-strand), while only 31% of the coding sequences are present on the leftward-transcribed strand (l-strand).

The positions of promoters and starts of transcription have been mapped via a variety of methods (Berk and Sharp, 1977b; Pettersson and Mathews, 1977; Spector *et al.*, 1978; Seghal *et al.*, 1979; Wilson *et al.*, 1979; Chow *et al.*, 1979a,b; Shaw and Ziff, 1980; Akusjärvi and Persson, 1981a; Stillman *et al.*, 1981). Many of the positions of promoters have been correlated with sequences generally indicated as TATA or Goldberg–Hogness boxes. These AT-rich sequences are considered to represent a constitutive part of promoter signals (see Chapter 5). The genes expressed early in infection are transcribed from six different promoters (r-strand: positions 1.3, 4.6, 16.5, and 76.6; l-strand: 75.1 and 99.1). The intermediate genes are transcribed from promoters located at positions 9.7 on the r-strand and 16.1 and 75.1 on the viral l-strand. The long late transcription unit uses the major late promoter at map position 16.5 on the viral r-strand. All primary transcription products of adenovirus DNA are processed in the nucleus before entering the cytoplasm. They are capped with ^{7me}G5′pppN at the 5′ end, and they are polyadenylated at the 3′ end. With one exception (polypeptide IX mRNA), all primary transcription products are processed into families of related mRNAs that share common 5′ and 3′ ends, but differ by alternative splicing (early

regions E1a, E1b, E2a, E3, and E4, intermediate regions E2b and IVa_2, and late regions L1, L2, L3, L4, and L5). It should be noted that in fact, analysis of the late transcription unit of adenovirus led to the original discovery of the phenomenon of RNA splicing. A detailed analysis of the transcription of the adenovirus genome is presented in Chapter 5. The organization of the transcriptional units of the adenovirus genome will now be described systematically from left to right. Since the organization of the Ad2 and Ad5 genomes has been investigated most extensively, these genomes are used for illustration.

The positions of major landmarks of the transcription units are indicated in Figs. 3–6 and Appendix B in the r- and l-strand sequences. It should be borne in mind that sequences of the r-strand of DNA are equivalent to RNA transcribed from the l-strand and that sequences of the l-strand of the genome are equivalent to mRNA transcribed from the r-strand. Unfortunately, the entire nucleotide sequences of Ad2 and Ad5 are not yet available, only a number of noncontiguous regions having been sequenced. Therefore, the numbering of the base pairs in Fig. 3–6 and Appendix B has not been added, but the sequence of each specific region starts from the left with base pair number 1.

A. Early Region E1 (1.3–11.2)

Early region E1 is transcribed from the leftmost part of the viral r-strand. It contains genes involved in cell transformation (Graham *et al.*, 1974a,b; van der Eb *et al.*, 1979) and regulation of transcription (Berk *et al.*, 1979; Jones and Shenk, 1979a; Nevins, 1981). The complete nucleotide sequence of this region has been established for human serotypes Ad2, Ad5, Ad7, and Ad12 (van Ormondt *et al.*, 1978, 1980a,b; Sugisaka *et al.*, 1980; Dijkema *et al.*, 1980a,b, 1981; Bos *et al.*, 1981; Kimura *et al.*, 1981; Gingeras *et al.*, 1982). The overall organization of this region appears to be very similar for the different serotypes (van Ormondt *et al.*, 1980b; Dijkema *et al.*, 1982). The region between 1.3 and 11.2 m.u. can be subdivided into three transcription units designated E1a, E1b, and region IX (Kitchingman *et al.*, 1977; Berk and Sharp, 1977a, 1978; Chow *et al.*, 1979a,b). The mRNAs derived from region E1 have been characterized by EM mapping, *in vitro* translation, and sequence analysis. It appears that all mRNAs except protein IX mRNA have a spliced structure and code for a variety of proteins, some of which are structurally related.

1. Early Region E1a (1.3–4.6)

Early region E1a is transcribed from the r-strand between 1.3 and 4.6 m.u. and codes for proteins that are involved in initiation of transformation (van der Eb *et al.*, 1979) and regulation of early gene expression

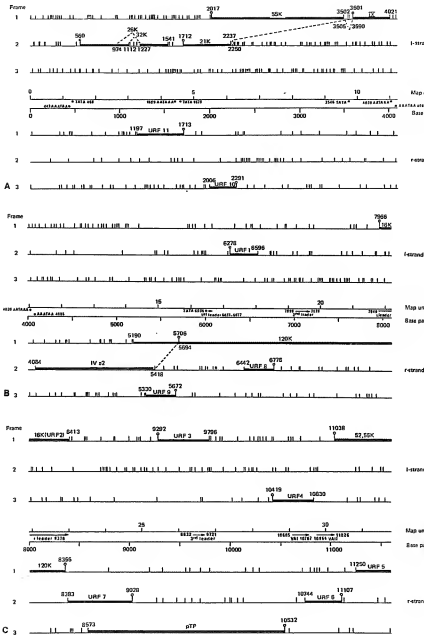

FIGURE 3A–C. Structural organization of the region between coordinates 0.0 and 31.7 on the Ad2 genome. The analysis of the structural organization is based on the nucleotide sequence shown in Fig. 18 [Appendix B], and indicated positions refer to this sequence. The l-strand of the DNA is homologous to r-strand transcripts, while the r-strand is homologous to l-strand transcripts. Here and in Figs. 4–6 and Appendix B: Termination codons [TAA, TGA, and TAG] are indicated in the three frames of the l- and r-strands by short vertical

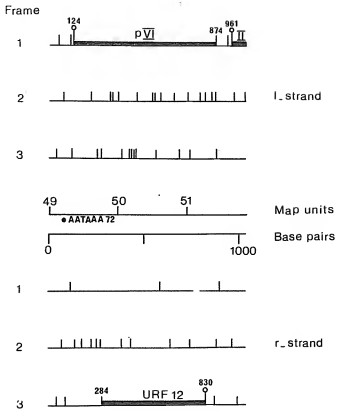

FIGURE 4. Structural organization of the region between coordinates 49.0 and 51.8 on the Ad2 genome. This analysis is based on the nucleotide sequence shown in Fig. 19 (Appendix B). This region mainly codes for the precursor of polypeptide VI. For explanation of the symbols, see the Fig. 3 caption.

(Jones and Shenk, 1979a; Berk *et al.*, 1979) (see Fig. 3). The promoter of this region has been mapped at position 1.3 (Wilson *et al.*, 1979). Analysis of the Ad2 sequence reveals that at position 468 [see Fig. 18 (Appendix B)], the TATA box TATTTATA is present. Baker and Ziff (1980, 1981) have characterized the position where transcription of the E1a RNA is initiated. They found that all mRNAs start with a capped dAMP residue

---

lines, while the initiation codon ATG is indicated by the symbol ♀. The coding regions that have been correlated with known proteins are shown by bold lines and are designated by molecular weights of the corresponding proteins or by roman numerals. Unidentified reading frames [(URF) initiating with ATG and terminating with one of the termination codons] or open reading frames [(ORF) regions between two termination codons] longer than 300 nucleotides are also indicated. Between the scales for Map units and Base pairs, the positions of TATA boxes, polyadenylation signals, and leader sequences are indicated. At some positions along the genome, splicing may occur. These positions are indicated by interrupted lines.

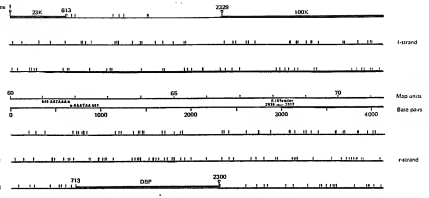

FIGURE 5. Structural organization of the region between coordinates 59.9 and 71.4 on the Ad5 genome. This analysis is based on the nucleotide sequence shown in Fig. 24 (Appendix B). This region codes for a 23K protein, DNA-binding protein (DBP), and a part of the 100K protein. For explanation of the symbols, see the Fig. 3 caption.

derived from position 499. Three mRNA species have been identified from region E1a with sedimentation coefficients of 13, 12, and 9 S. These mRNAs share the same 5′ and 3′ termini and differ only in the size of the RNA fragment removed by splicing during the processing of nuclear RNA (Kitchingman *et al.*, 1977; Berk and Sharp, 1977a, 1978; Chow *et al.*, 1979a,b; Perricaudet *et al.*, 1979). The splice points of the 13 S RNA have been mapped at nucleotide positions 1112 and 1227 and of the 12 S mRNA at positions 974 and 1227 (Perricaudet *et al.*, 1979). The donor splice site of the 9 S mRNA species has not been determined yet. The 3′ ends of the mRNAs are located at nucleotide position 1630, while the polyadenylation signal AATAAA is found at position 1609 (Perricaudet *et al.*, 1979; Fraser *et al.*, 1982).

Since the reading frames in the E1a mRNAs are the same, the proteins derived from these mRNAs share their N-terminal and C-terminal segments and differ only in the number of intervening amino acids. From the DNA sequence, the complete amino acid sequences of the proteins specified by the 13 and 12 S mRNA species can be predicted. Both proteins must be rich in Pro and Glu residues and have theoretical molecular weights of 32 and 26K, respectively. The protein derived from the 9 S mRNA has an estimated molecular weight of 13K. These proteins have been correlated with proteins produced during cell-free translation of isolated mRNAs (Lewis *et al.*, 1976; Pettersson and Mathews, 1977; Harter and Lewis, 1978; Green *et al.*, 1979a; Esche *et al.*, 1980; Spector *et al.*, 1980a,b; van der Eb *et al.*, 1979; Lupker *et al.*, 1980). These translation products with apparent molecular weights of 48–58, 42–54, and 28K are structurally related, which is in agreement with the nucleotide sequence of this region. The discrepancy between the theoretical and apparent molecular weights probably reflects the extremely high Pro contents of these proteins, which lead to aberrant migration in gels.

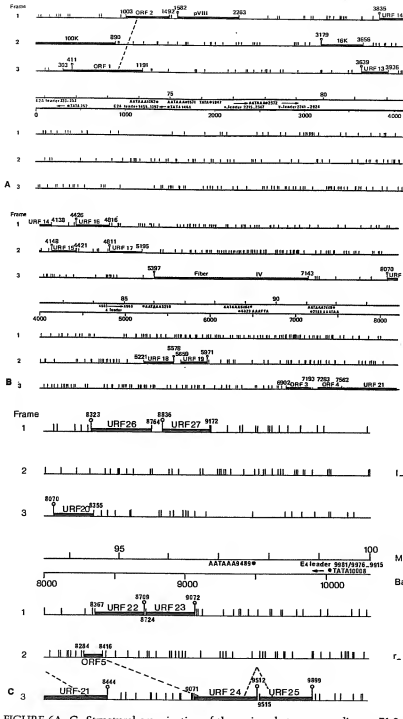

FIGURE 6A–C. Structural organization of the regions between coordinates 71.2 and 100.0 on the Ad2 genome. This analysis is based on the nucleotide sequence shown in Fig. 21 [Appendix B]. For explanation of the symbols, see the Fig. 3 caption.

As mentioned before, the E1a regions of Ad2, Ad5, Ad7, and Ad12 show very similar organization. In all serotypes, three spliced mRNA species are synthesized. Recently, it was shown that the protein encoded by the 13 S mRNA governs early gene expression (Montell *et al.*, 1982).

2. Early Region E1b (4.6–11.2)

Early region E1b is transcribed from the viral r-strand between map coordinates 4.6 and 11.2 [see Figs. 3 and 18 (Appendix B)]. The proteins encoded by this region are involved in transformation and play an important role in oncogenesis; during lytic infection, these proteins are involved in DNA replication (Harrison *et al.*, 1977; Frost and Williams, 1978; Jones and Shenk, 1979a,b; van der Eb *et al.*, 1979; Bernards *et al.*, 1982; van den Elsen *et al.*, 1982). Little is known about the precise role of these proteins. Studies of cells transformed by DNA fragments of different length have suggested that region E1a is able to immortalize cells, while region E1b is required for full expression of the typical phenotype of adenovirus-transformed cells (van der Eb *et al.*, 1979; Houweling *et al.*, 1980).

The promoter of early region E1b is located at map position 4.6, where, at nucleotide 1670, a Goldberg–Hogness box TATATAA is found (Fig. 18). Transcription may start at position 1700 or 1702 (Baker and Ziff, 1981) and proceeds until nucleotide 4061 (Perricaudet *et al.*, 1980; Fraser *et al.*, 1982). The polyadenylation signal of region E1b is located at nucleotide 4030. The primary transcription product of region E1b is processed by splicing into a 22 and a 13 S mRNA species. Both species share a 3′-terminal segment from nucleotide 3590 to a polyadenylation site at nucleotide 4061. Both species also contain a 5′-terminal sequence from 1700 or 1702 to a donor splice site at nucleotide 2250. In the 13 S mRNA, nucleotide 2250 is joined to an acceptor splice site at 3590, whereas the 22 S mRNA includes nucleotide 2250 to a second donor splice site at nucleotide 3505. Nucleotide 3505 of the 22 S mRNA is ligated to the common acceptor splice site at nucleotide 3590. From these points, the mRNA sequence continues to the polyadenylation site near nucleotide 4061 (Perricaudet *et al.*, 1980; Alestrom *et al.*, 1980). *In vitro* translation experiments have shown that two major proteins with molecular weights of 55–65 and 15–19K can be assigned to this transcription unit (Lewis *et al.*, 1976; Harter and Lewis, 1978; van der Eb *et al.*, 1979; Brackmann *et al.*, 1980). This observation is in agreement with the fact that the two mRNA species contain information for two major tumor (T) antigens with theoretical molecular weights of 21 and 55K, which are encoded by two overlapping reading frames. The 22 S mRNA codes for both proteins depending on which particular ATG triplet serves as the start codon. The 21K protein initiates at the 5′-proximal ATG (position 1712), while the 55K protein initiates at the second ATG (nucleotide 2017) in another reading frame (Anderson and Lewis, 1980; Bos *et al.*, 1981). In addition,

the 21K protein can also be synthesized from the 13 S mRNA. Peptide mapping has shown that the small-t and the large-T antigens do not share tryptic peptides, in accordance with the nucleic acid sequence data (Bos *et al.*, 1981).

Similar organization of region E1b has been found for Ad2, Ad7, and Ad12 (Bos *et al.*, 1981; Kimura *et al.*, 1981; Dijkema *et al.*, 1982; Gingeras *et al.*, 1982). This does not exclude small differences between mRNAs from different serotypes. Comparison of the E1b mRNAs of Ad5 and Ad12 has revealed that the Ad12 mRNA contains additional splices in the 3' noncoding part of the mRNA (Virtanen *et al.*, 1982a). The precise functions of the 21K and 55K proteins are still unknown.

The 22 and 13 S mRNAs both contain information for protein IX, a protein that has been mapped between 9.7 and 11.2 map units (Chow *et al.*, 1977b; Pettersson and Mathews, 1977; Esche *et al.*, 1980). However, this information is not translated from these messengers. Instead, a unique short mRNA is synthesized from an independent transcription unit between coordinates 9.7 and 11.2 (Wilson *et al.*, 1979; Chow *et al.*, 1977a,b; Pettersson and Mathews, 1977). The sequences of the genes that encode the Ad2 and Ad5 polypeptides IX have been established, which allowed the identification of transcription and translation signals (Maat *et al.*, 1980; Alestrom *et al.*, 1980). The polypeptide IX TATA box is located at position 3546, and transcription starts at nucleotide position 3575 or 3577 (map position 9.7) in the Ad2 sequence [Fig. 18 (Appendix B)]. Its 3' end has been located at nucleotide position 4061 (map position 11.2) (Alestrom *et al.*, 1980; Fraser *et al.*, 1982), while the polyadenylation signal AATAAA is located at position 4030. The same polyadenylation signal is also used for processing of the large and the small E1b T antigen mRNAs. The RNA synthesized is not processed and represents the only known unspliced adenovirus mRNA. The mRNA contains a continuous open reading frame that codes for a protein of 14K. Protein IX (apparent molecular weight 12.5K) is found in virions and was therefore originally classified as a late protein (Pettersson and Mathews, 1977). Later experiments showed that protein IX is also synthesized in the absence of viral DNA replication, indicating that it is an intermediate protein (Persson *et al.*, 1978). The complete nucleotide sequence of the polypeptide IX gene has been determined for human serotypes Ad2, Ad3, Ad5, Ad7, and Ad12 (Maat *et al.*, 1980; Alestrom *et al.*, 1980; Dijkema *et al.*, 1981; Kimura *et al.*, 1981; Engler, 1981). Within the same group, the protein IX genes exhibit a striking similarity, but the genes of serotypes from different groups are much less homologous.

3. Unidentified Reading Frames

In the l-strand transcripts, a number of unidentified reading frames (URFs) have been detected. The URFs larger than 300 nucleotides are indicated in Figs. 3 and 18 (Appendix B). However, recently it could be

shown that in transformed cells and infected cells, an l-strand transcript is synthesized that spans the E1a–E1b junction and codes for a protein with a molecular weight of 11K (Katze *et al.*, personal communication). This transcript might very well be derived from URF 11 located between nucleotides 1713 and 1197 on the viral l-strand. At position 443, the sequence AATAAA is found, which might function as a polyadenylation signal. This indicates that it is certainly not impossible that later some of these will appear to be expressed during the infection cycle, albeit at a very low frequency.

B. Late and Intermediate Genes in the Region between Coordinates 11.2 and 31

1. Major Late Promoter and Tripartite Leader

The region between 11.2 and 31 contains a mosaic of different strategic regions in both complementary strands [see (Figs. 3 and 18 (Appendix B)]. The major late promoter has been mapped on the r-strand at position 16.5 (Evans *et al.*, 1977; Ziff and Evans, 1978). This promoter is also active early in infection (Shaw and Ziff, 1980; Akusjärvi and Persson, 1981b). In the nucleotide sequence at this position, there is a TATA box TATAAA at nucleotide position 6006, and transcription starts from position 6037 (Baker and Ziff, 1981). During early times in infection, transcription proceeds no further than map position 39, while at late times, transcription proceeds to map position 99.0 (Fraser *et al.*, 1979). Messenger RNAs derived from r-strand transcripts starting at position 16.5 contain a common tripartite leader (Berget *et al.*, 1977, 1978; Chow *et al.*, 1977a,b; Akusjärvi and Pettersson, 1979a,b; Zain *et al.*, 1979a,b; Ziff and Evans, 1978). The sequence of the tripartite leader of late Ad2 RNA has been determined by sequencing complementary DNA (cDNA) transcribed from hexon mRNA and a cDNA clone of fiber mRNA (Zain *et al.*, 1979a; Akusjärvi and Pettersson, 1979b). The tripartite leader sequences have been established for a number of serotypes [Ad2 (Ziff and Evans, 1978; Akusjärvi and Pettersson, 1979a; Zain *et al.*, 1979a), Ad5 (van Beveren *et al.*, 1981), Ad3 and Ad7 (Engler *et al.*, 1981)].

The overall length of the Ad2 tripartite leader is 203 nucleotides, comprising 41 nucleotides from the promoter region at map position 16.5, 72 nucleotides from position 19.6, and 90 nucleotides from position 26.5 on the genome. Examination of the sequence reveals that the tripartite leader does not contain an AUG triplet, suggesting that translation of late adenoviral mRNA does not initiate within the tripartite leader. In some intermediate and late transcripts, an additional leader fragment (i-leader) has been detected by R-loop mapping, which maps at coordinates 21.5–23.0 (Chow *et al.*, 1979a). Sequence analysis has shown that in contrast to the tripartite leader, the i-leader (nucleotides 7940–8379) contains an open reading frame for a hypothetical protein of 15.9 kilodaltons (kd).

In vitro translation of mRNA selected on DNA fragments that contain i-leader sequences does indeed lead to synthesis of a hitherto unknown protein (URF2) with an apparent molecular weight of 13.6–16K (Lewis *et al.*, 1979; Lewis and Mathews, 1980; Virtanen *et al.*, 1982b). The termination codon for the 15.9-kd protein is not present in the i-leader, but is probably located within the third leader. The function of the 15.9-kd protein is still unknown.

2. Virus-Associated RNAs

At positions 28.8 and 29.5 on the genome, the genetic information for two low-molecular-weight RNAs is located, these RNAs being designated virus-associated (VA) RNAs VA-RNAI and VA-RNAII (Söderland *et al.*, 1976; Mathews and Pettersson, 1978) (Fig. 3). In contrast to all other genes, the VA genes are transcribed by RNA polymerase III instead of RNA polymerase II (Price and Penman, 1972; Weinman *et al.*, 1974, 1976; Söderland *et al.*, 1976). The VA-RNAs are probably synthesized from two separate promoter sites in the r-strand and do not undergo post-transcriptional processing. The genes and the RNA products have been subjected to nucleotide sequence analysis (Ohe and Weissman, 1970, 1971; Ohe, 1972; Pan *et al.*, 1977; Celma *et al.*, 1977a,b; Akusjärvi *et al.*, 1980). The nucleotide sequence of VA-RNAI was determined by Ohe and Weissman (1971) to be 157–160 nucleotides long (nucleotides 10,608–10,764/10,767). Vennström *et al.* (1978a,b) demonstrated that the 5′ end of VA-RNAI is heterogeneous and may start at nucleotide 10,605 or 10,608 [Fig. 18 (Appendix B)]. The length of VA-RNAII is 158–163 nucleotides (nucleotides 10,864–11,021/11,026), and the two VA-RNAs are separated by a spacer about 98 nucleotides long. The function of these RNAs is still unknown; so far, no proteins derived from them have been found. It has been suggested that these RNAs play a role in splicing or stabilization of late mRNA (Murray and Holliday, 1979; Mathews, 1980). It is interesting to note that the VA-RNAs can form almost identical secondary structures with high stability. The structures show similarities to transfer RNA (Zain *et al.*, 1979b; Akusjärvi *et al.*, 1980).

3. Early Region E2b and Protein IVa_2 (11.2–30.2)

For a long time, it has been thought that the l-strand transcripts between map units 11 and 30 coded only for the intermediate protein IVa_2 (molecular weight 50K), a protein that is involved in the morphogenesis of virions (Persson *et al.*, 1979a). The gene of this protein has been mapped between coordinates 11.3 and 16.1 (Lewis *et al.*, 1975, 1977) [see Figs. 3 and 18 (Appendix B)]. Transcription of the IVa_2 gene starts from a promoter located at map position 16.1. Nucleotide sequences of this region reveal that although no regular TATA box is located in this region, the sequence TCCTT, which may resemble a TATA box, is pres-

ent at nucleotide 5859. RNA synthesis starts at position 5826 or 5824 and proceeds to nucleotide 4051 (Alestrom *et al.*, 1980; Baker and Ziff, 1981; Fraser *et al.*, 1982) (Fig. 18). The messengers from this region contain an intron located between nucleotides 5419 and 5693 (Chow *et al.*, 1977a,b; Broker *et al.*, 1977; Kilpatrick *et al.*, 1979; van Beveren *et al.*, 1981). The mRNA contains a long open reading frame (ORF) corresponding to 445 amino acids of which the first 4 N-terminal amino acids are coded by RNA upstream from the donor splice site and the remaining amino acid residues by RNA downstream from the acceptor splice site. It is noteworthy that the reading frame in which these 4 N-terminal amino acids lie is part of a much longer reading frame that codes for a protein of 120 kd (see below). Another interesting feature of the IVa_2 gene is that the 3' end of the message overlaps the end of the E1b and polypeptide IX mRNAs with 9 nucleotides. Also, the IVa_2 termination codon TAA (nucleotide 4084) forms a part of the IVa_2 polyadenylation signal AATAAA (nucleotide 4086). The IVa_2 genes of serotypes Ad2, Ad5, and Ad7 have all been sequenced and show the same structural organization (van Beveren *et al.*, 1981; Engler and van Bree, 1982; Gingeras *et al.*, 1982; Alestrom *et al.*, 1982b). The IVa_2 nucleotide sequences of Ad7 and Ad5 are 78% homologous.

A new class of mRNAs from the region between 11 and 30 m.u. was identified by Stillman *et al.* (1981). The promoter of these transcripts has been mapped at position 75.1 and is probably identical to the promoter of early region E2a. Transcripts of this region, which is designated E2b, contain, in addition to the 75.1-m.u. leader, additional leaders from 68.5 and 39 m.u. Region E2b has been classified as an intermediate transcription unit (Fig. 3). The main bodies of messages derived from this transcription unit may start at positions 30, 26, and 23, respectively, and continue to position 11.2. *In vitro* translation of preselected mRNAs derived from the region between 11.2 and 31.5 led to synthesis of proteins with molecular weights of 105, 87, and 75K (Stillman *et al.*, 1981; Binger *et al.*, 1982). The 87K protein is identical to the precursor terminal protein (pTP) with a molecular weight of 80K described by Challberg *et al.* (1980) (see Section VI). Nucleotide sequence analysis of this region has indicated the presence of two long ORFs located between 28.9 and 23.5 m.u. and 24.1 and 14.2 m.u. [Fig. 18 (Appendix B)]. The region between 28.9 and 23.5 m.u. beginning at nucleotide 10,577 has the first ATG at nucleotide 10,532 and continues to a terminator at nucleotide 8573. This frame codes for a protein with a minimum molecular weight of 74.5K. The second large ORF begins at nucleotide 8793, has the first ATG at 8355, and continues to a terminator TAG at nucleotide 5190. The total coding capacity of this reading frame is 132.1kd, while the capacity from the first ATG to the terminator is 120.4kd (Gingeras *et al.*, 1982; Alestrom *et al.*, 1982; Engler *et al.*, 1983). Since the precise structure of the spliced E2b mRNAs is still unknown, it cannot be excluded that a part of the leader from map position 39 is part of the coding sequences of E2b mRNAs. EM

mapping of E2b mRNAs has indicated that the 3′ ends of the messengers map at position 11.2, the same position where the 3′ end of IVa_2 mRNA is located. It is therefore likely that the mRNAs of pTP and the 120kd polypeptide have the same 3′ end and polyadenylation site as the IVa_2 mRNA (Aleström *et al.*, 1980; Stillman *et al.*, 1981). Smart and Stillman (1982) showed by analysis of tryptic peptides from the terminal protein and its precursor that the ORF between 28.9 and 23.5 codes for pTP. Very recently, the ORF from 24.1 to 14.2 was assigned to an adenovirus-specific DNA polymerase (Kelly, Stillman, and Hurwitz, personal communications). This polymerase has an apparent molecular weight of 140K, co-purifies with pTP, and is able to complement a defective *in vitro* DNA replication system of the DNA-synthesis-negative temperature-sensitive (*ts*) mutant Ad5*ts*36 (Enomoto *et al.*, 1981; Lichy *et al.*, 1982; Kelly and Stillman, personal communications). The mutant Ad5*ts*36 has been mapped between 18.5 and 22.0 m.u. (Galos *et al.*, 1979). In addition to these two proteins, all E2b messengers contain genetic information for the IVa_2 protein, but this information is probably not translated from the E2b messengers.

4. Unidentified Reading Frames

Several unidentified shorter reading frames are present in this region of the viral genome (Fig. 3). However, no correlation with known proteins or gene functions has been discovered yet. In this respect, it should be noted that translation *in vitro* of early mRNA selected by hybridization to fragments of DNA derived from this region has identified mRNA species that encode additional proteins (Lewis and Mathews, 1980). A DNA fragment from 17.0 to 21.5 m.u. selects an mRNA that is complementary to the r-strand and codes for a 13.5-kd protein (Lewis *et al.*, 1979; Lewis and Mathews, 1980). Further, two polypeptides of 16.5 and 17.0kd have been described, translated from mRNAs that are selected by DNA fragments lying between 11.6 and 17.0 m.u. (Lewis *et al.*, 1979).

C. Late Regions L1, L2, and L3 (31.0–61.7)

A major event in the infection cycle of adenoviruses is the activation of the entire late transcription unit. As mentioned in Section VII.B.1, the promoter of the late transcription unit is located at map position 16.5, and this promoter is already active early in infection. However, during the early phase, transcription does not proceed further than map position 39 (Shaw and Ziff, 1980; Akusjärvi and Persson, 1981b). In the late phase, transcription continues to map position 99.0 (Fraser *et al.*, 1979, 1982). The transcription product ranging from map positions 16.5 to 99.0 is considerably processed, leading to the production of five families of late

mRNAs (L1–L5) (Chow *et al.*, 1977b; McGrogan and Raskas, 1978; Chow and Broker, 1978; Nevins and Darnell, 1978). Each of the five classes expresses more than one protein and contains mRNAs with a common 3′ end (Ziff and Fraser, 1978; Nevins and Darnell, 1978; Fraser and Ziff, 1978). At the 5′ end, all these mRNAs contain the tripartite leader.

The region on the Ad2 genome between 30.2 and 61.7 m.u. contains the genes for the families L1–L3. As mentioned above, the L1 family of RNAs is already expressed early in infection. This family consists of three mRNAs that have a common 3′ end mapping at 39 m.u. At the same position, the polyadenylation site of the L1 family has been mapped (Fraser *et al.*, 1979, 1982). The L1 mRNAs code for two structurally related proteins of 52 and 55K (Lewis and Mathews, 1980; Miller *et al.*, 1980) and polypeptide IIIa (molecular weight 66K). Since nucleotide sequences from the left-hand end of Ad2 DNA have not been established further than position 31.5, only the initiation codon of the 52,55K protein has been identified unambiguously (Akusjärvi *et al.*, 1980). The function of the 52,55K protein is still unknown. The L1 family further contains genetic information for protein IIIa, which has been mapped by hybrid-arrest translation between 34.3 and 39.3 m.u. This protein has a molecular weight of 66K and is present in virions associated with the hexon polypeptides.

Located from positions 39 to 50 is the L2 family, consisting of three mRNA species that code for polypeptide III (molecular weight 85K), the precursor of polypeptide VII (20K), and polypeptide V (48.5K). These proteins are all constituents of adenovirus particles. One of these, the precursor of polypeptide VII, is processed during maturation of virions to mature polypeptide VII (molecular weight 18.5K). This protein is identical to the major core protein. The genes for protein III, the precursor of protein VII, and protein V have been mapped by R-loop mapping and hybrid-arrest translation at 37.4–43.9, 43.9–45.4, and 45.3–49.6, respectively (Miller *et al.*, 1980).

Fraser *et al.* (1982) have mapped the polyadenylation site of the L2 family at position 50. This fits well with the fact that in the nucleotide sequence from the region between coordinates 49.0 and 51.8 [Fig. 19 (Appendix B)], the polyadenylation site of the L2 family has been identified at nucleotide 92, while an AATAAA signal is present at nucleotide 72 (Akusjärvi and Persson, 1981a).

The nucleotide sequence data from region 49.0–51.8 make it possible to pinpoint exactly some landmarks of the L3 family of late mRNAs (see Figs. 4 and 19). Three species of mRNAs have been identified that can be translated into the precursor of polypeptide VI (pVI), hexon (polypeptide II), and a 23K protein. The gene for polypeptide pVI is located from 49.1 to 51.2 and has been sequenced completely (Miller *et al.*, 1980; Akusjärvi and Persson, 1981a). Also, the acceptor splice site at which the 5′ leader sequences are joined to the pVI message has been determined (nucleotide 123) (Fig. 19). This splice site is situated very close to the

start codon (nucleotide 124). The gene for polypeptide pVI codes for a protein with a theoretical molecular weight of 27K. This protein is cleaved during maturation of young virions, resulting in the formation of polypeptide VI (molecular weight 24K), which is part of the adenovirion. With the help of nucleotide sequence analysis, the N-terminal end of the hexon polypeptide has been mapped at coordinate 51.6, while the C terminus is located at 59.7 (Akusjärvi and Pettersson, 1978a,b). The hexon polypeptide is translated from start codon 961 of an mRNA that contains, in addition to the tripartite leader, a main body starting at nucleotide 925 in the sequence of Fig. 19 (Appendix B) to nucleotide 836 in the sequence of Fig. 20.1. The common polyadenylation site of the L3 RNAs has been mapped at the same position. In accord with other polyadenylation sites, the sequence AATAAA is located close to this addition site (nucleotide 812) (Fig. 20.1). The total nucleotide sequence of the hexon gene has not been established yet; only stretches of nucleotides have been determined (Jörnvall *et al.*, 1981b). However, by combination of nucleotide sequence and amino acid sequence data, the complete amino acid sequence of the Ad2 hexon polypeptide has been established (Jörnvall *et al.*, 1981a). It appears that the hexon polypeptide of Ad2 consists of 966 amino acid residues. It is the largest viral protein and has a calculated molecular weight of 108K and an apparent molecular weight of 120K.

From positions 59.9 to 61.7, r-strand transcripts code for a protein of molecular weight 23K (Kruijer *et al.*, 1980; Akusjärvi *et al.*, 1981) [see Figs. 5 and 20.2 (Appendix B)]. A minor RNA species consisting of the tripartite leader and a main body corresponding to this region has been identified and translated. A protein with a molecular weight of 23K is synthesized from this messenger. Since the Ad2 mutant *ts*1 has been mapped in the L3 region and is hampered in proteolytic cleavage of precursors of polypeptides VI, VII, and VIII, it has been suggested that the 23K protein is identical to a virus-coded protease (Bhatti and Weber, 1979).

D. Early Region E2a (61.5–75.1)

Early region E2a codes for the single-strand-specific, DNA-binding protein (DBP) (Figs. 5 and 6). This protein, discovered by van der Vliet and Levine (1973), is phosphorylated, has an apparent molecular weight of 72K, and is involved in DNA replication, in regulation of early and late gene expression, and in cell transformation (Ginsberg *et al.*, 1974; van der Vliet *et al.*, 1975, 1977; van der Vliet and Sussenbach, 1975; Carter and Ginsberg, 1976; Horwitz, 1978; Mayer and Ginsberg, 1977; Carter and Blanton, 1978; Nevins and Jensen-Winkler, 1980; Klessig and Grodzicker, 1979). The DBP genes of Ad2 and Ad5 have been analyzed in most detail. Therefore, the positions of strategic signals in the DBP gene are described in these sequences [Figs. 21 and 24 (Appendix B)]. It should be

pointed out that the main bodies of the Ad2 and Ad5 genes are highly homologous. The promoter for region E2a is located at 75.1 m.u. on the viral l-strand and is used early in infection (Baker *et al.*, 1979). At this position, the sequence TCCTTAA (nucleotide 1490) (Fig. 2.1) is found, which is an aberrant type of TATA box. This promoter is probably also used for transcription of the E2b transcription unit. At later times in infection, transcription of the E2a region starts from a promoter at map position 72.0, where the TATA box TACAAATTT is found (nucleotide 352) (Fig. 2.1). A minor start of transcription at intermediate and late times is found at 87 m.u. Recently, an additional promoter sequence was identified about 26 nucleotides upstream from the major early promoter (Mathis *et al.*, 1981). The function of the minor promoter sequence is still unknown.

Depending on the time in infection, mRNA species from the E2a region contain two different short leaders. Depending on the time post-infection, one is derived from position 75.1 [nucleotides 1392–1458/1459 (67/68 nucleotides long)] or 72.0 (nucleotides 253–321/323 (69/71 nucleotides long)] [Fig. 21 (Appendix B)]. The other is derived from position 68.8 [nucleotides 2936–3012 (77 nucleotides long)] (Baker *et al.*, 1979; Kruijer *et al.*, 1981, 1983) [Fig. 24 (Appendix B)]. The main body of the E2a mRNAs is located between map positions 66.5 and 61.5 [Fig. 24 (nucleotides 2309–642)] (Kruijer *et al.*, 1981; Akusjärvi *et al.*, 1981). The site of polyadenylation has been localized at nucleotide 642, while the sequence AATAAA is found at position 661 (Akusjärvi *et al.*, 1981; Fraser *et al.*, 1982). From the nucleotide sequence of the E2a region and the structure of DBP mRNAs, it can be derived that all coding sequences of these RNAs are located within the main body (Kruijer *et al.*, 1981, 1982). Translation starts at ATG 2300 and runs to stop codon 713. The Ad2 and Ad5 mRNAs code for a protein of 529 amino acids (molecular weight 59K), while Ad12 DBP is 484 amino acid residues long (molecular weight 54K). Comparison of the Ad2 and Ad5 DBP nucleotide sequences reveals a high degree of homology, with only 9 amino acid differences in the corresponding amino acid sequences. However, Ad5 and Ad12 DBPs differ considerably in nucleotide and amino acid sequences. These differences are mainly located in the N-terminal part of the DBP molecule. In contrast, the C-terminal regions of the DBP molecules show a high degree of homology (80%) (Kruijer *et al.*, 1983). It is especially this part of the molecule that is involved in DNA replication (Ariga *et al.*, 1980; Kruijer *et al.*, 1981). The terminal part of DBP is involved in regulation of late expression (Klessig and Grodzicker, 1979; Kruijer *et al.*, 1981).

E. Late Region L4 (66.5–77.3)

This region includes a set of r-strand transcripts that code for a 100-kd protein (66.5–73.1), a 33-kd protein (71.5–74.0), and the precursor of

polypeptide VIII (molecular weight 26K) (75.5–77.3) (Figs. 5 and 6). The indicated map positions have been determined by hybrid-arrest translation (Miller *et al.*, 1980). Polypeptide VIII (molecular weight 13K) is produced by proteolytic cleavage of its precursor during maturation of virions and is in virions associated with the hexon capsomers. The 100-kd protein is involved with folding of the hexon polypeptide chains into trimers (Ginsberg, personal communication), while the function of the 33-kd protein is still unknown. The four mRNAs that code for these proteins form the L4 family of late mRNAs and share the 3′-terminal sequences. The common polyadenylation site has been mapped at 78 map units.

Nucleotide sequences of this region have been determined in Ad2 and Ad5 DNA (Galibert *et al.*, 1979; Hérissé *et al.*, 1980; Kruijer *et al.*, 1981, 1982). Therefore, the strategic landmarks of the L4 proteins can be indicated at the nucleotide level. The acceptor splice point of the Ad5 100-kd polypeptide has been determined by reverse transcription of 100-kd mRNA and is located at nucleotide 2316 [Fig. 24 (Appendix B)] (Kruijer *et al.*, 1983). The polyadenylation site of the L4 mRNAs is mapped close to the sequence AATAAA at nucleotide 2572 [Fig. 21 (Appendix B)] (Fraser *et al.*, 1982). Comparison of the Ad5 sequence, which extends to coordinate 71.4, with the sequence of Ad2 indicates that nucleotides 3855–4107 of the Ad5 sequence (Fig. 24) are colinear with nucleotides 1–253 of the Ad2 sequence (Fig. 21). The frames in the overlapping sequences are identical and code, with a single exception, for identical amino acids. Using the combined sequences, it is possible to construct a hybrid 100-kd protein consisting of an amino-terminal part from Ad5 and a carboxy-terminal part of Ad2. The hypothetical hybrid protein consists of 805 amino acids and has an actual molecular weight of 89K.

The coding sequences of the 100 and 33-kd proteins partially overlap. However, since these proteins do not share tryptic peptides (Gambke and Deppert, 1981), it is most likely that they are encoded by r-strand transcripts in different ORFs. While the information for the 100-kd protein terminates at nucleotide 890, two ORFs (ORFs 1 and 2) can be distinguished in the other two reading frames, viz., ORF 1 from nucleotides 306 to 1191 (between stop codons 303 and 1191) and ORF 2 from nucleotides 1006 to 1492 (between stop codons 1003 and 1492 (Fig. 21). An ATG is present at nucleotide 411. Since one of the L4 mRNAs contains an internal splice that maps reasonably well in the region where these two ORFs overlap, it is likely that these regions code for the 33-kd protein. However, this has still to be proved by experimental data. One of the three short additional leaders for the fiber mRNA (x-leader) is also transcribed in this region from the r-strand (77.2–77.6). The x-leader has not been sequenced yet, but employing EM mapping data and typical RNA splice-site sequences, it has been inferred that this leader is transcribed from the r-strand from nucleotides 2215 to 2347. The l-strand between 66.5 and 77.3 units codes for the DBP mRNA leaders from positions 75.1,

72.0, and 68.8, respectively. The structure of the corresponding TATA boxes and individual leaders was described in Section VII.D.

F. Early Region E3 (76.6–86.0)

This region, located between coordinates 76.6 and 86.0, codes for a large number of r-strand transcripts and polypeptides (Fig. 6). At least six major species of mRNAs have been identified, coding for proteins of 13, 14, 15.5–16, and 19–21 kd, respectively (Lewis *et al.*, 1976; Harter *et al.*, 1976; Green *et al.*, 1979d; Ross *et al.*, 1980). The polypeptides of 19–21 kd are glycoproteins, which are associated with the membrane fraction (Persson *et al.*, 1979b, 1980a). Tryptic peptide analysis has shown that the 16-kd polypeptide is the unglycosylated precursor of the 19-kd protein (Persson *et al.*, 1980b).

The mRNAs from this region share sequences at their 5′ ends from coordinates 76.6 to 77.6, which are ligated to sequences starting at 78.6 m.u. The 3′ ends of the transcripts may vary.

Nucleotide sequence analysis of this region has revealed that a TATA box of the structure TATAA is located at nucleotide 1947 (76.7 m.u.), while transcription starts at nucleotide 1976/1978 (Baker and Ziff, 1981) [Fig. 21 (Appendix B)]. In region E3, two polyadenylation sites are present, one of which has been mapped at the nucleotide level (nucleotide 4148). Examination of the sequence of this region reveals that the sequence ATTAAA is found at position 4136. This sequence differs from the common hexanucleotide AATAAA that is found in all other Ad2 mRNAs associated with the polyadenylation site. In the sequence of region E3, the sequence AATAAA is located at nucleotide 5209, which fits very well with EM mapping data of some E3 mRNA species. However, for these messengers, the polyadenylation site has not yet been determined in detail.

The first ATG in the E3 region is found at position 2266, which suggests that E3 mRNAs have a 290-nucleotide-long untranslatable region at their 5′ ends. About 80 nucleotides downstream from this ATG lies a potential splice site, and this site fits very well with the position where the common leader sequence of E3 mRNAs has been mapped (positions 76.6–77.6). This leader sequence may code for 27 amino acid residues, which would be common to all E3 proteins. However, determination of the amino-terminal sequence of the unglycosylated 16-kd protein has shown that translation of the coding sequence of this protein starts at nucleotide 3179 and continues to nucleotide 3656. This codes for a protein of 159 amino acids with a molecular weight of 18.4K. Obviously, the ATG at position 2266 present in all E3 mRNAs is not recognized during translation. If the 3′ splice point of the first E3 intervening sequence is located around position 2840 (Hérissé *et al.*, 1980), this implies that the mRNA for the 16-kd protein has an untranslated region

more than 700 nucleotides long. Region E3 contains a number of short URFs. A hypothetical organization of translation is indicated in Fig. 6. Unfortunately, no data are available to assign the URFs unambiguously to individual proteins. As described above, the only exception is the 16-kd protein. The function of the E3 proteins is completely obscure. In some adenovirus–simian virus 40 hybrids, this region is absent without affecting the viability of the virus. Apparently this region is nonessential for viral multiplication (for a review, see Tooze, 1981). In addition to the E3 proteins, this region codes for two additional leaders of the fiber mRNAs, viz., the y-leader (78.6–79.2) and the z-leader (84.7–85.1) (Chow and Broker, 1978). Only the y-leader has been sequenced and appears to be located at nucleotides 2741–2924 (Zain *et al.*, 1979a). Employing EM mapping data and the common sequences of RNA splice sites, it has been inferred that the z-leader is located at nucleotides 4805–4963 (Hérissé *et al.*, 1980).

G. Late Region L5 (86.0–91.3)

The L5 family of late transcripts consists of two major mRNA species that code for a single virion protein, the fiber (polypéptide IV). The main bodies of these RNAs map between coordinates 86.0 and 91.3 (Miller *et al.*, 1980) (Fig. 6). RNA from this region differs from all other late messengers in that it may contain, in addition to the common tripartite leader, additional leader sequences (x, y, and z) from map positions 77.2, 78.6, and 84.7 (Chow and Broker, 1978; Zain *et al.*, 1979a). The y-leader is the most abundant additional leader of fiber mRNA; however, even this leader is not present in all RNA species. It has been shown that the presence or absence of the y-leader does not influence the translation of fiber mRNA. Even in the absence of the y-leader, the mRNA can be translated normally to fiber protein in an *in vitro* translation system (Dunn *et al.*, 1978). The nucleotide sequence of this leader has been established to be 184 nucleotides long, and although an ATG is present in this sequence, it is obviously not employed and not required for appropriate translation of fiber mRNA.

The complete nucleotide sequence of region L5 has been established (Zain *et al.*, 1979a; Zain and Roberts, 1979; Hérissé and Galibert, 1981; Hérissé *et al.*, 1981; Gingeras *et al.*, 1982) [Fig. 21 (Appendix B)]. The 5′ end of the main body of the fiber mRNA is located at nucleotide 5395, adjacent to the codon of fiber mRNA at position 5397 (Zain and Roberts, 1979; Zain *et al.*, 1979a). The termination codon of the fiber gene is located at nucleotide 7143 and is part of the polyadenylation signal AATAAA at position 7141. The mRNA codes for 582 amino acid residues that contitute a protein with a theoretical molecular weight of 61.9K, which agrees very well with the apparent molecular weight of the fiber protein of 62K.

H. Early Region E4 (91.3–99.2)

Early region E4 messengers are transcribed from the viral l-strand between coordinates 91.3 and 99.0 and code for a large set of polypeptides (Fig. 6). The promoter of this region has been mapped at 99.2 m.u., while the 3′ ends of E4 RNAs have been localized at 91.3 m.u. (Berk and Sharp, 1978; Chow *et al.*, 1979a,b; Baker and Ziff, 1981; Hashimoto *et al.*, 1981).

All E4 mRNAs share their 5′- and 3′-terminal nucleotide sequences, but vary in the location of splice points (Berk and sharp, 1978; Chow *et al.*, 1979a; Kitchingman and Westphal, 1980). These messengers code for a number of polypeptides with molecular weights of 11, 13, 17, 19, 21, and 24K (Lewis *et al.*, 1976; Green *et al.*, 1979d; Ross *et al.*, 1980). As yet, these proteins have not been assigned unambiguously to individual mRNA species. Only the position of the acidic 11K polypeptide has been correlated to a specific region in the nucleotide sequence of this region (Hérissé *et al.*, 1981).

Besides the fact that the synthesis of the E4 proteins starts about 2 hr after infection, reaches a maximum around 3 hr, and then declines, these proteins seem to be nonessential for DNA replication, and their role is at present unknown.

Recently, the complete Ad2 nucleotide sequence of this region has been established (Shinagawa *et al.*, 1980; Hérissé *et al.*, 1981; Gingeras *et al.*, 1982) [Fig. 21 (Appendix B)], while for Ad5, the region between 97 and 100 m.u. has been determined (Steenbergh and Sussenbach, 1979) [Fig. 25.1 (Appendix B)]. At nucleotide 10,008 in the Ad2 sequence, a TATA box with the structure TATATATA can be recognized as part of a promoter sequence. Transcription begins with the sequence TTTTTA at nucleotides 9981–9976, leading to a heterogeneous array of starts (Baker and Ziff, 1981) (Fig. 21). All major species of mRNAs contain a leader sequence starting at the cap sites and probably terminating at nucleotide 9915, where a potential 5′ splice site is located. This leader sequence is devoid of ATG able to play a role in initiation of translation. Therefore, such a signal should be located in the body of the various mRNA species spliced to this leader sequence. At the other end of the sequence, transcription terminates close to an AATAAA sequence, which is located at position 7188. This is consistent with EM mapping data of E4 RNAs. It should be pointed out that transcription sometimes proceeds beyond this point to coordinate 61.5, leading to the production of a minor species of E2a mRNA (see Fig. 1).

The nucleotide sequence of the E4 region reveals that a large number of short URFs are present in all three reading frames.

Comparison of the nucleotide sequence and the mRNA mapping data indicates that there is a reasonably good correlation between the mapping data and potential donor and acceptor splice sites in the sequence. From the predicted structure of the various spliced mRNA species, a hypo-

thetical translation pattern has been proposed (Hérissé *et al.*, 1981; Gingeras *et al.*, 1982). However, only in the case of the acidic 11K protein could its coding region be deduced with reasonable certainty from the nucleotide sequence to be located in URF 23. Further nucleotide sequence analysis of mRNAs and translation of individual mRNA species is required to determine unambiguously the relationship between individual RNAs and the corresponding proteins.

I. Unidentified Reading Frames

In addition to the URFs of early region E4, an additional ORF with a coding capacity of 12kd (ORF 3) is found in the viral l-strand transcripts (Fig. 6). This region is located between stop codons at positions 7193 and 6902 and starts with AAA (7190) (Fig. 21). At nucleotide 7166, the first ATG codon is found, while at nucleotide 6323, even the sequence ATTAAA is present, which resembles an aberrant type of polyadenylation signal also present in early region E3. It should be noted that although the major E4 transcription termination site has been mapped at 91.3 m.u., Nevins *et al.* (1980) have calculated that transcription termination takes place at 88.4 m.u., which corresponds very well with the sequence ATTAAA at nucleotide 6323 (Hérissé *et al.*, 1981). However, no mRNA species derived from this region are currently known. The same holds for two URFs in r-strand transcripts that code for proteins with theoretical molecular weights of 10.6 and 12K (URFs 26 and 27).

VIII. COMPARISON OF GENOMES AND CONCLUDING REMARKS

The organization of the adenovirus genome as described in Section VII has mainly been restricted to Ad2 because the most detailed information is available for this serotype. However, it should be emphasized that for all serotypes the structure of which has been investigated, the same overall organization has been observed. For a number of serotypes, nucleotide sequence data are available. These data are compiled in Appendix B, including the analysis of these sequences. For a number of genes, the nucleotide sequences have been compared, as well as the amino acid sequences of the corresponding proteins. Van Ormondt *et al.* (1980b) have analyzed the homology among the E1a regions of Ad5, Ad7, and Ad12, while Bos *et al.* (1981) and Kimura *et al.* (1981) have studied the homology of the E1b regions of Ad5 and Ad12. The IVa_2 and polypeptide IX genes of Ad2, Ad3, Ad5, and Ad7 have been compared (Dijkema *et al.*, 1981; Engler, 1981; Engler and van Bree, 1982), as well as the late leaders of Ad2, Ad3, and Ad7 (Engler *et al.*, 1981) and the E2b regions of Ad2 and Ad7 (Engler *et al.*, 1983). The redundancies of different serotypes were

analyzed by Tolun *et al.* (1979) and Shinagawa and Padmanabhan (1980), while the DNA-binding protein genes of Ad2, Ad5, and Ad12 were compared by Kruijer *et al.* (1981, 1982, 1983).

Detailed analysis of the organization of the adenovirus genome reveals that the available coding information of this virus is used in a very economical fashion. Unraveling of the information at the nucleotide level reveals all kinds of peculiar properties in its organization. There are spliced and unspliced mRNA species (e.g., hexon and polypeptide IX RNA), overlapping termination codons and AATAAA signals (e.g., fiber and IVa_2 RNA), overlapping genes (e.g., the 33- and 100-kd proteins), and symmetrical transcription (120-kd protein and the 16-kd i-leader product). There are classic TATA boxes (e.g., E1a proteins) and polyadenylation signals (AATAAA) (hexon RNA) and aberrant sequences with the same function (TATA box TCCTT (E2a early promoter) and polyadenylation signal ATTAAA (region E3)).

In conclusion, the adenovirus genome is a microuniverse in itself, and the study of its organization and regulation of expression is a great joy and satisfaction for every scientist who dedicates herself or himself to the unraveling of its secrets.

ACKNOWLEDGMENTS. The author gratefully acknowledges the very valuable assistance of Mr. O. van Hien for providing computer facilities and Dr. T. Broker for maps and other information. Without their help, this chapter would never have been completed. He also thanks M. M. Kwant, M. G. ter Braak-Kuijk, W. van Driel, F. M. A. van Schaik, E. Simon, W. Kruijer, A. W. M. Rijnders, J. van der Rijst, and H. Laanen for technical assistance and Dr. P. C. van der Vliet for critical reading of the manuscript. He gratefully acknowledges the fact that his colleagues Drs. J. Engler, R. J. Roberts, K. Fujinaga, M. Horwitz, U. Pettersson, H. van Ormondt, R. Padmanabhan, B. Stillman, E. Ziff, and F. Galibert have made available new data prior to publication.

APPENDIX A: RESTRICTION ENDONUCLEASE CLEAVAGE MAPS

This appendix contains a compilation of restriction maps of the genomes of different adenovirus serotypes (Figs. 7–17). These maps have partially been published and partially been presented as personal communications. Most of these maps have been compiled before by Tooze (1981) and are redrawn with permission from the Cold Spring Harbor Laboratory Publication Department. The coordinates of the Ad1, Ad2, and Ad5 maps have been recalculated (Gingeras *et al.*, 1982). Details on the origin of the maps are indicated in Tooze (1981), unless otherwise stated.

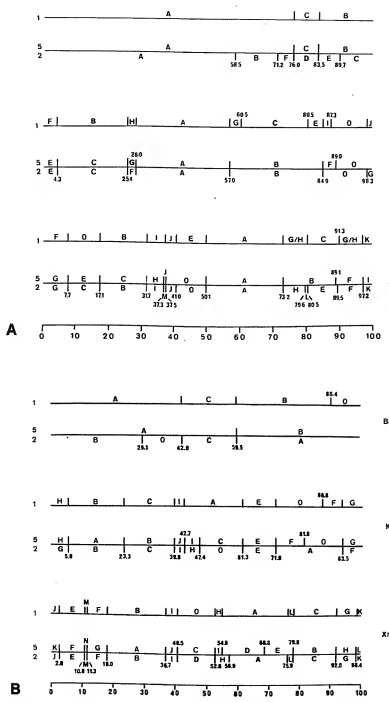

FIGURE 7A–D. Restriction endonuclease cleavage maps of Group C Ad1, Ad2, and Ad5.

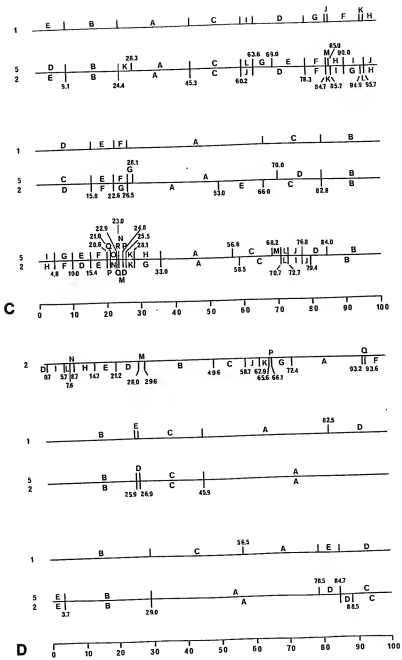

FIGURE 7 (*Continued*)

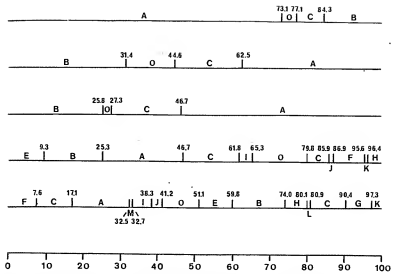

FIGURE 8. Restriction endonuclease cleavage maps of Group C Ad6. The maps were determined by Naroditsky *et al.* (1980) and oriented such that the transforming region is located at the left. The *Eco*RI map was determined by Forsblom *et al.* (1976).

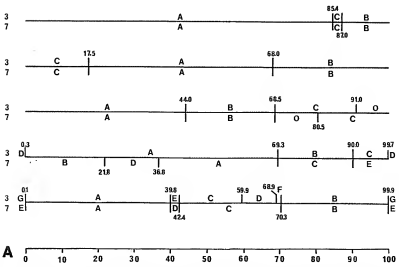

FIGURE 9A, B. Restriction endonuclease cleavage maps of Group B Ad3 and Ad7. The *Bst*EII and *Bcl*I maps were determined by R. Padmanabhan (personal communication).

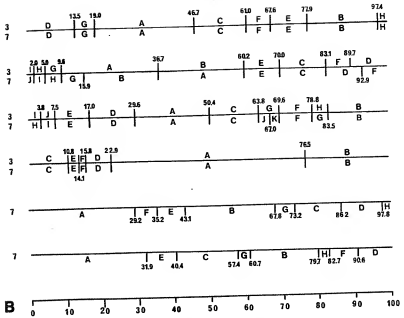

FIGURE 9 (*Continued*)

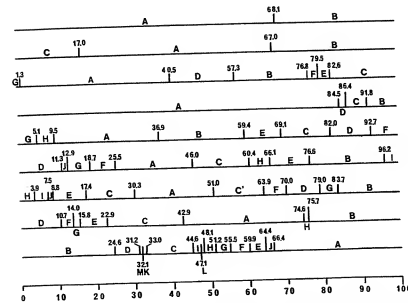

FIGURE 10. Restriction endonuclease cleavage maps of Group B Ad16 (Chang 79). These maps were determined by Varsanyi *et al.* (1977), Winberg and Hammarskjöld (1980), and Hammarskjöld and Winberg (personal communication).

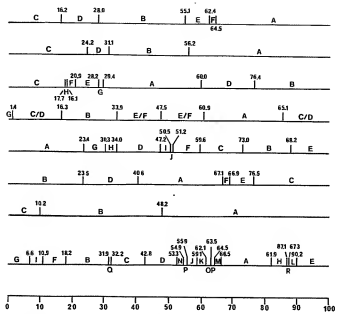

FIGURE 11. Restriction endonuclease cleavage maps of Group A Ad12 (Huie).

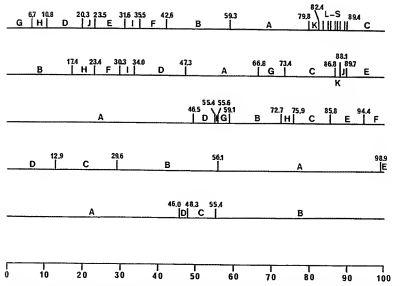

FIGURE 12. Restriction endonuclease cleavage maps of Group A Ad31 (strain 1315). The maps were determined by Y. Sawada, Y. Yamashita, F. Kamda, K. Sekikawa, and K. Fujinaga (personal communication).

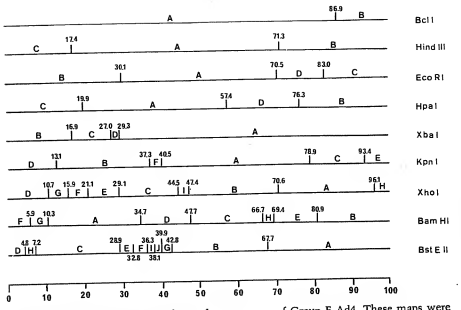

FIGURE 13. Restriction endonuclease cleavage maps of Group E Ad4. These maps were determined by Tokunaga *et al.*.(1982).

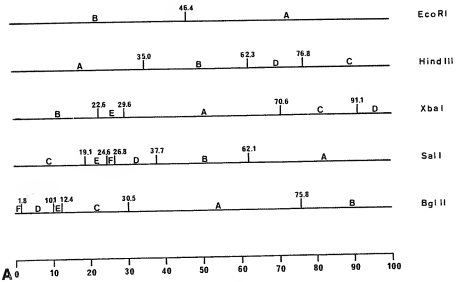

FIGURE 14A, B. Restriction endonuclease cleavage maps of simian adenovirus type 7. The *Eco*RI, *Sal*I, and *Bg*II maps of simian adenovirus [strain C8] were determined by Naroditsky *et al.* [1980] and oriented with respect to the conventional genetic map by Ponomareva *et al.* [1979], who located the transforming region to the left. The other maps were determined by T. I. Tikchonenko and colleagues [personal communication].

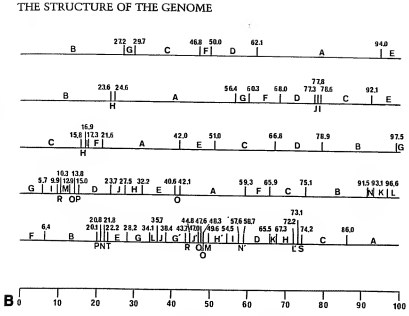

FIGURE 14 (*Continued*)

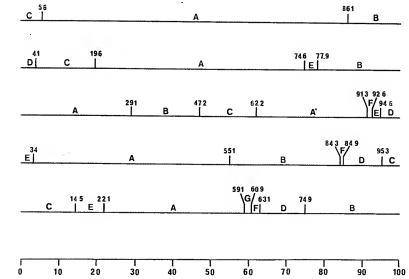

FIGURE 15. Restriction endonuclease cleavage maps of simian adenovirus type 20. These maps were determined by T. I. Tikchonenko and colleagues (personal communication).

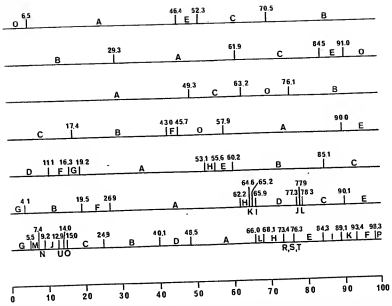

FIGURE 16. Restriction endonuclease cleavage maps of simian adenovirus type 30. The *Eco*RI and *Bgl*II maps were determined by Dimitrov *et al.* (1979). They were originally reported to be those of simian adenovirus type 38, and identification subsequently revised by Tikchonenko and colleagues (personal communication), who also determined the other maps.

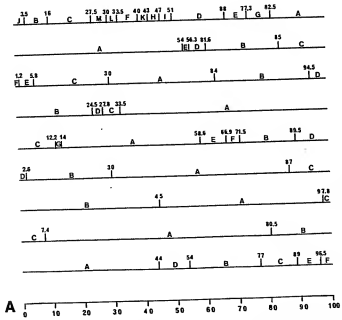

FIGURE 17A, B. Restriction endonuclease cleavage maps of mouse adenovirus type FL. These maps were determined by Larsen *et al.* (1979). For the orientation, see Larsen *et al.* (1979) and Temple *et al.* (1981).

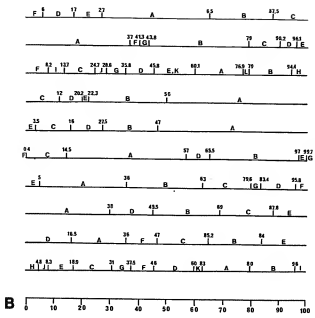

FIGURE 17 (*Continued*)

APPENDIX B: NUCLEOTIDE SEQUENCES

This appendix contains a compilation of nucleotide sequences partially published and partially presented as personal communications (Figs. 18–29). Since r-strand transcripts are homologous to the l-strand, the positions of important landmarks for r-strand transcripts are indicated in the l-strand sequence. Likewise, strategic sequences for l-strand transcripts are indicated in the r-strand sequence. The sequences of Ad2 and Ad5 are very homologous. Therefore, it has been supposed that specific signals identified in the sequence of one serotype also indicate the positions of these signals in the sequence of the other serotype. The positions of the inverted terminal repetition boundaries and start and termination codons of known coding regions are indicated, as well as the positions of 5′ and 3′ ends of mRNAs, splice points, and TATA boxes. The latter signals are supposed to be a constitutive part of transcriptional promoters. The sequences AATAAA and ATTAAA, which are found within about 30 nucleotides from the 3′ end of the mRNAs, are underlined. These sequences have been associated with polyadenylation. Open reading frames (ORFs), defined as regions between two termination codons in the same frame, have been indicated when the size exceeds 300 nucleotides. The same holds for unidentified reading frames (URFs) (regions that start with an ATG codon and terminate with one of the termination codons).

FIGURE 18A–L. Nucleotide sequence of a region between coordinates 0.0 and 31.7 on Ad2 DNA. This sequence was determined by Gingeras *et al.* (1982b) and Alestrŏm *et al.* [1982] [nucleotides 5776–11,558]. Although base pair 9 (AT) is absent in the sequence determined by Gingeras and co-workers, most other investigators do find this base pair in their Ad2 by Gingeras and co-workers, most other investigators do find this base pair in their Ad2 by strains. In the latter case, the terminal sequences of Ad2 and Ad5 are identical. To allow comparison of Ad2 and Ad5 sequences, this base pair was included in the sequence presented, which leads to a numbering one nucleotide higher than that of Gingeras and co-workers. Short stretches of this sequence were previously determined by Arrand and Roberts [1979], Shinagawa and Padmanabhan [1979], Perricaudet *et al.* [1979], Alestrŏm *et al.* [1980], and Baker and Ziff [1980]. The identification of strategic sequences is indicated in Sections VII.A and VII.B.

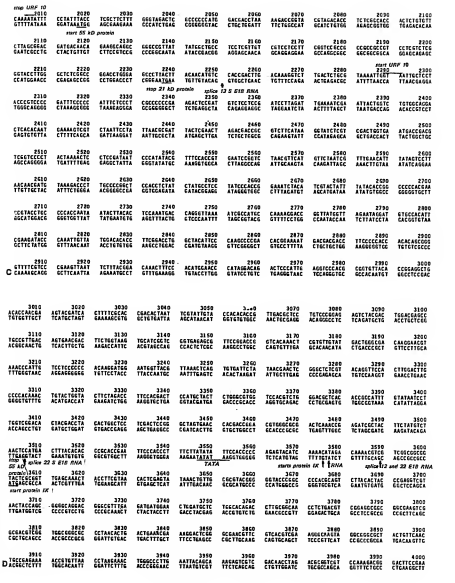

FIGURE 18 (*Continued*)

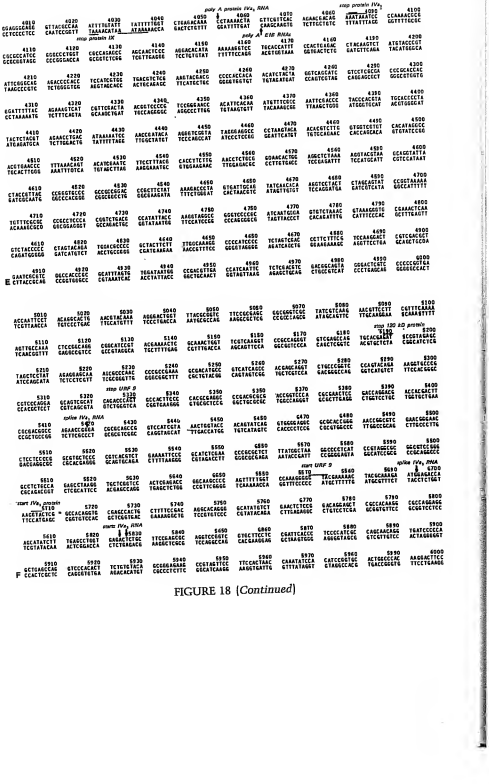

FIGURE 18 (*Continued*)

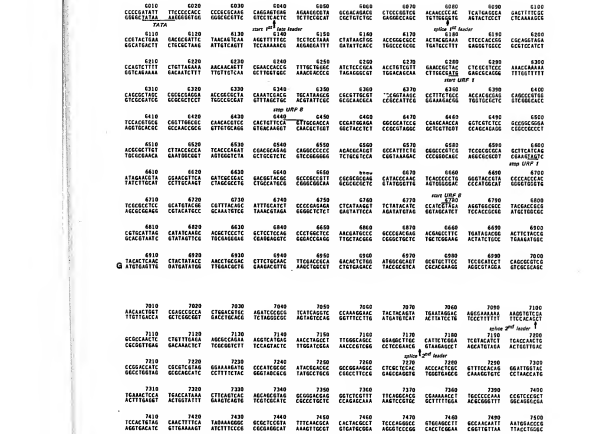

FIGURE 18 (*Continued*)

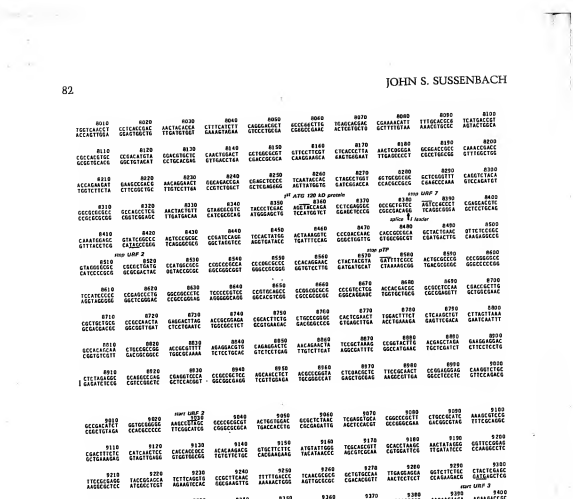

FIGURE 18 (*Continued*)

FIGURE 19. Nucleotide sequence of a region between coordinates 49.0 and 51.8 on the Ad2 genome. This sequence was determined by Akusjärvi and Persson (1981a). The positions of strategic signals were determined by Akusjärvi and Pettersson (1979a) and Akusjärvi and Persson (1981a).

FIGURE 18 (*Continued*)

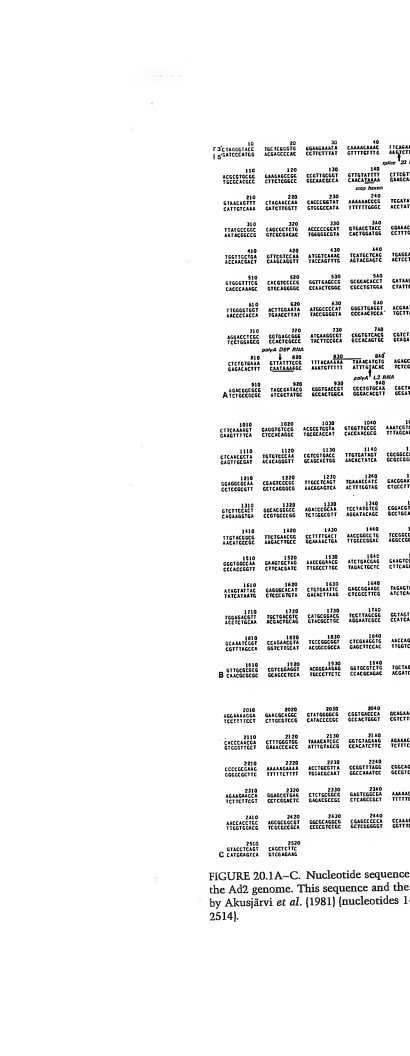

FIGURE 20.1A–C. Nucleotide sequence of a region between coordinates 59.5 and 66.4 on the Ad2 genome. This sequence and the positions of strategic sequences were determined by Akusjärvi *et al.* (1981) (nucleotides 1–1164) and Kruijer *et al.* (1982) (nucleotides 858–2514).

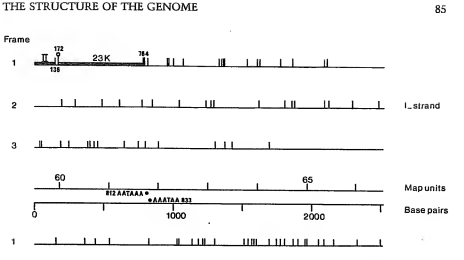

FIGURE 20.2. Structural organization of a region between coordinates 59.5 and 66.4 on the Ad2 genome. This map is derived from the nucleotide sequence in Fig. 20.1. For explanation of the symbols, see the Fig. 3 caption (Section VII).

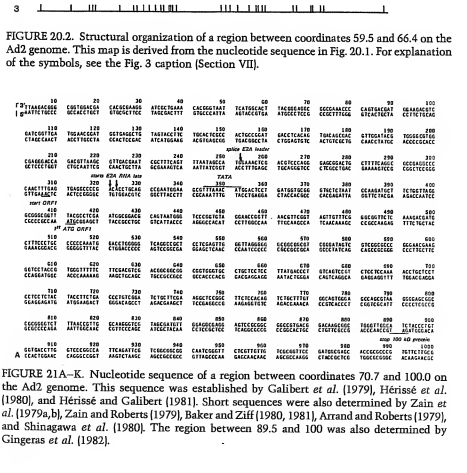

FIGURE 21A–K. Nucleotide sequence of a region between coordinates 70.7 and 100.0 on the Ad2 genome. This sequence was established by Galibert *et al.* (1979), Hérissé *et al.* (1980), and Hérissé and Galibert (1981). Short sequences were also determined by Zain *et al.* (1979a,b), Zain and Roberts (1979), Baker and Ziff (1980, 1981), Arrand and Roberts (1979), and Shinagawa *et al.* (1980). The region between 89.5 and 100 was also determined by Gingeras *et al.* (1982).

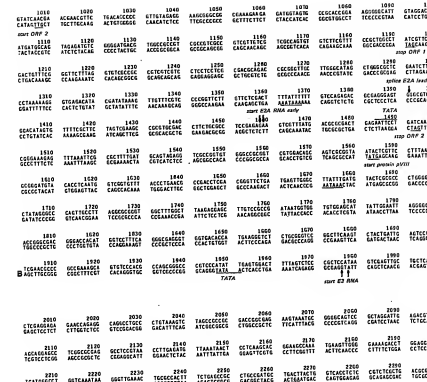

FIGURE 21 (*Continued*)

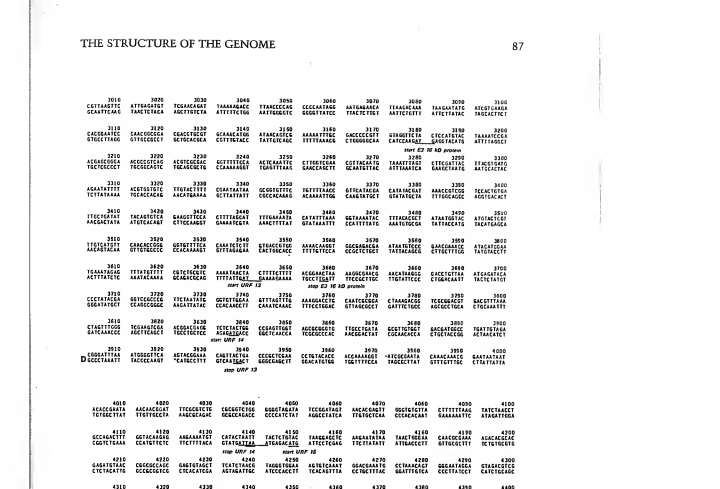

FIGURE 2.1 (*Continued*)

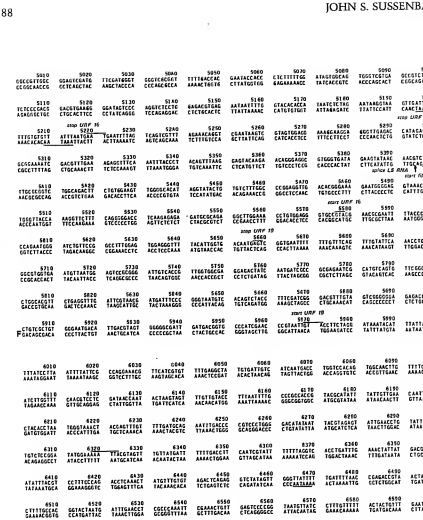

FIGURE 21 (*Continued*)

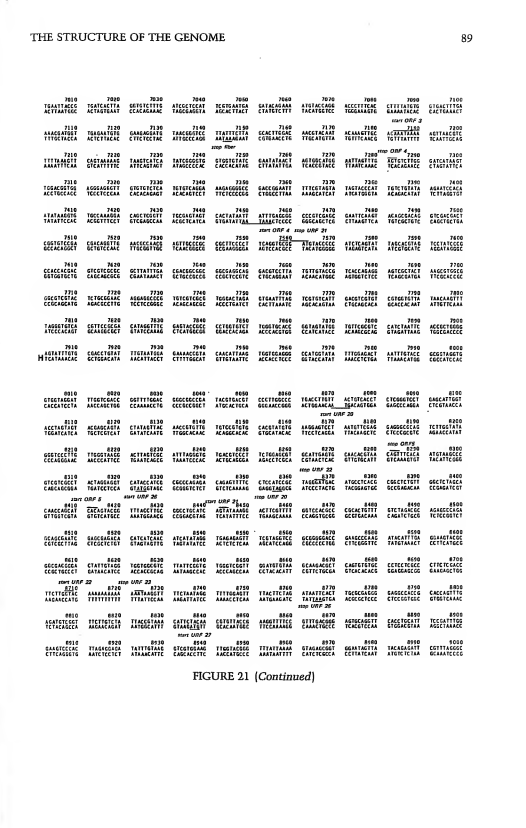

FIGURE 21 (*Continued*)

FIGURE 21 (*Continued*)

FIGURE 22. Nucleotide sequence of a region between coordinates 9.6 and 11.2 on the Ad3 genome. This sequence was established by Engler (1981). The region codes for polypeptide IX. For the positioning of strategic signals, see Fig. 3 (Section VII).

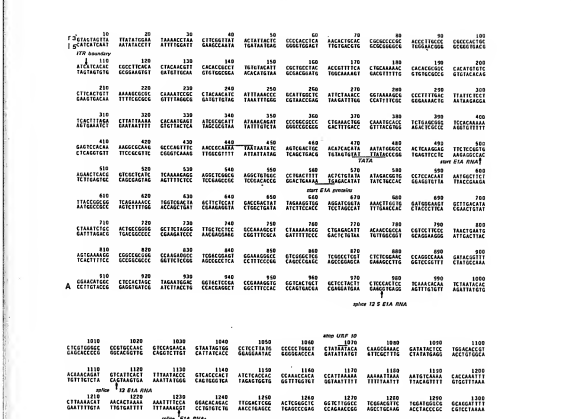

FIGURE 23.1A–L. Nucleotide sequence of a region between coordinates 0.0 and 31.7 on the Ad5 genome. This sequence was established by Steenbergh *et al.* (1977), van Ormondt *et al.* (1978), Maat and van Ormondt (1979), Maat *et al.* (1980), van Beveren *et al.* (1981), Bos *et al.* (1981), and H. van Ormondt and B. M. M. Dekker (personal communication). For interpretations, see van der Eb *et al.* (1979) and van Ormondt *et al.* (1980a,b).

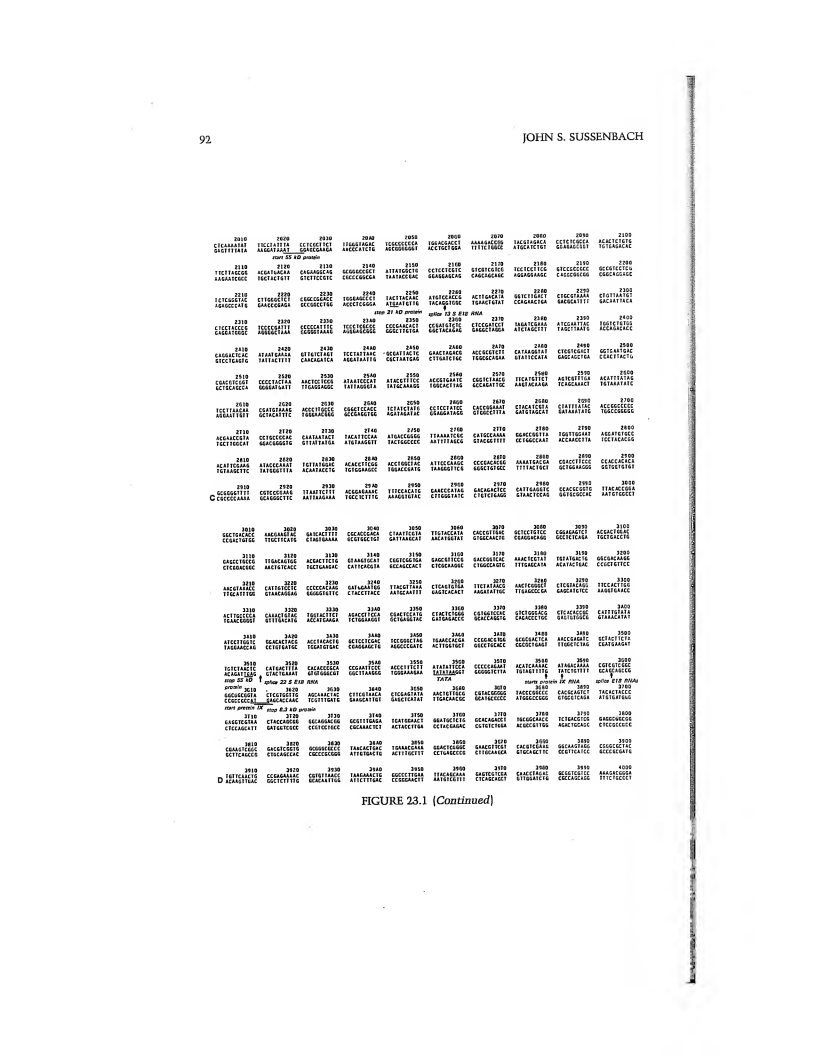

FIGURE 23.1 (*Continued*)

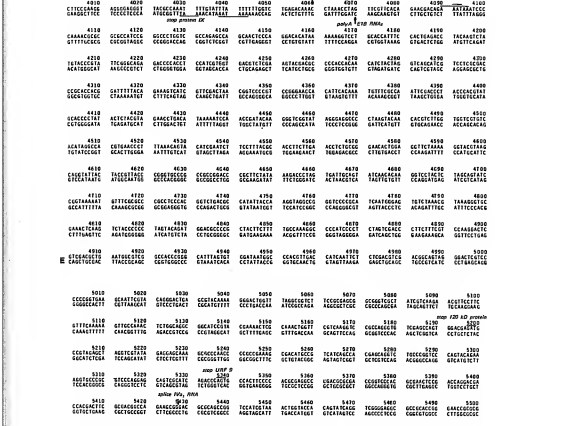

FIGURE 23.1 (*Continued*)

FIGURE 23.1 (*Continued*)

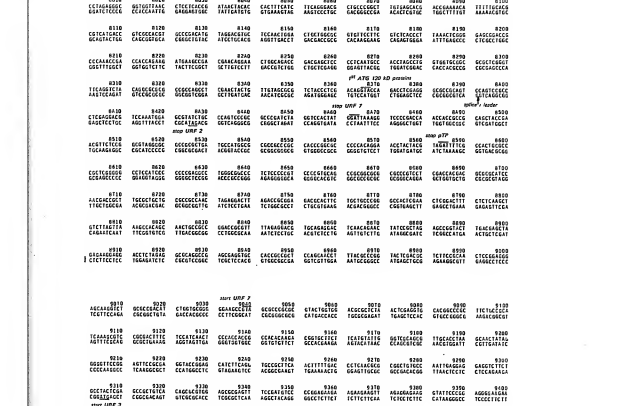

FIGURE 23.1 (*Continued*)

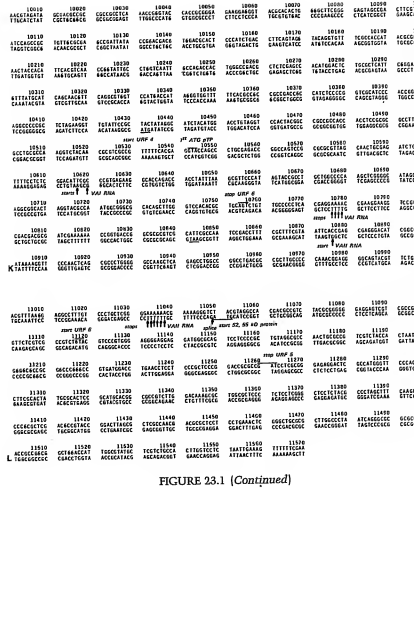

FIGURE 23.1 (*Continued*)

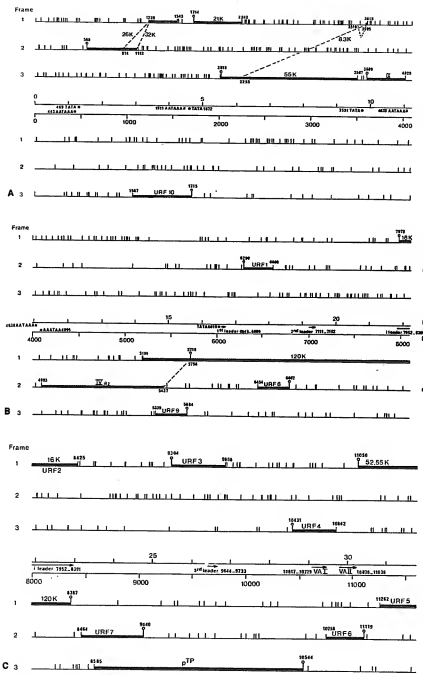

FIGURE 23.2A–C. Structural organization of a region between coordinates 0.0 and 31.7 on the Ad5 genome. This map is derived from the nucleotide sequence in Fig. 23.1. For the positioning of strategic signals, see Fig. 3 (Section VII).

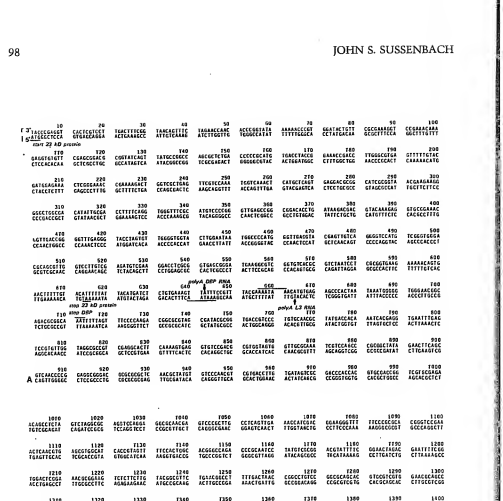

FIGURE 24A–D. Nucleotide sequence of a region between coordinates 59.9 and 71.4 on the Ad5 genome. This sequence and the positions of splice points and leaders were determined by Kruijer *et al.* (1980, 1981, 1983). A schematic presentation of this sequence is shown in Fig. 5 [Section VII].

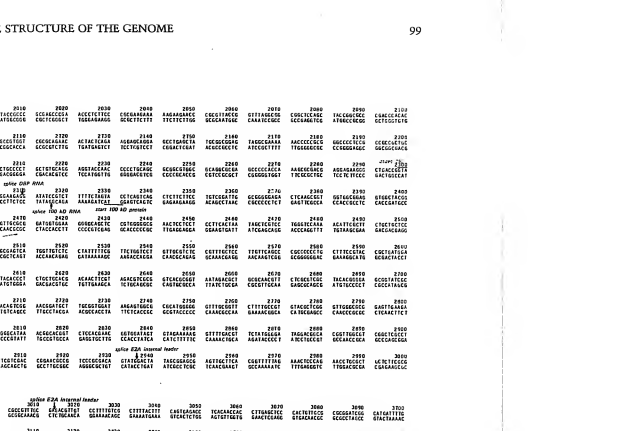

FIGURE 24 (*Continued*)

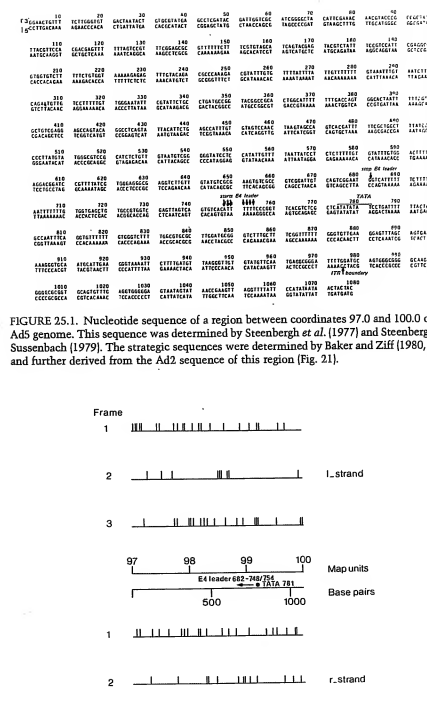

FIGURE 25.1. Nucleotide sequence of a region between coordinates 97.0 and 100.0 on the Ad5 genome. This sequence was determined by Steenbergh *et al.* [1977] and Steenbergh and Sussenbach [1979]. The strategic sequences were determined by Baker and Ziff (1980, 1981) and further derived from the Ad2 sequence of this region (Fig. 21).

FIGURE 25.2. Structural organization of a region between coordinates 97.0 and 100.0 on the Ad5 genome. This map is derived from the nucleotide sequence in Fig. 25.1.

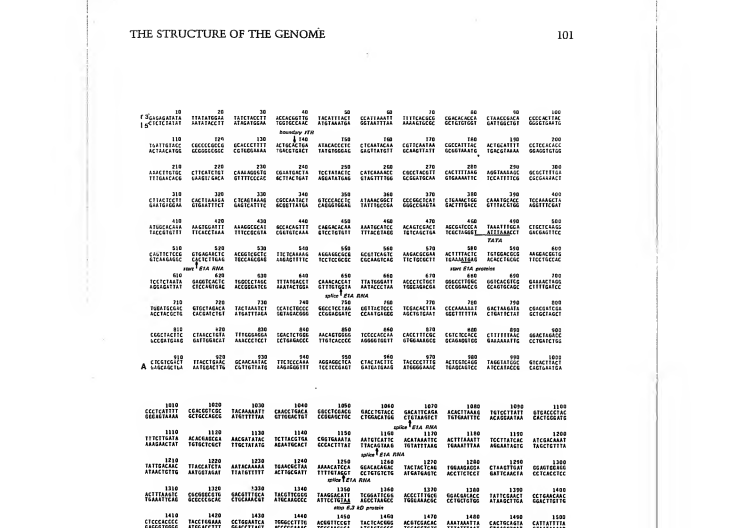

FIGURE 26.1A–K. Nucleotide sequence of a region between coordinates 0 and 31.7 on the Ad7 genome. This sequence and the positions of strategic sequences were established by Dijkema and Dekker (1979), Dijkema *et al.* (1980a,b, 1981, 1982), van Beveren *et al.* (1981), Engler (1981), Engler *et al.* (1981, 1983), and Engler and van Bree (1982).

```
      2010       2020       2030       2040       2050       2060       2070       2080       2090       2100
TAGAATCTAA TGACCGGTCA CGTCGGAGAC CCTCATCGTC CCTATGACTC TGTGGGTGGC TGGTACGGTC GCCAAGACGT CCTCCTCGTC GTCCTCCTGT
ATCTTAGATT ACTGGCCAGT GCAGCCTCTG GGAGTAGCAG GGATACTGAG ACACCCACCG ACCATGCCAG CGGTTCTGCA GGAGGAGCAG CAGGAGGACA

      2110       2120       2130       2140       2150       2160       2170       2180       2190       2200
TAGGCTCTCG GCCGGACCTG GGAGGCCACC TCCTCATCGA CTGGACAAAG GACTTGACGC TGCCCACGAA TGATCCAGAT GCTGGTCACC TGTCTTGTCC
ATCCGAGAGC CGGCCTGGAC CCTCCGGTGG AGGAGTAGCT GACCTGTTTC CTGAACTGCG ACGGGTGCTT ACTAGGTCTA CGACCAGTGG ACAGAACAGG
                                 stop 21 kD protein               splice↑E1B RNA

      2210       2220       2230       2240       2250       2260       2270       2280       2290       2300
CCTTAATTCT CCCTCTCCTT AGGATCACCC TTATTAAGTT CTTGGCTCAA CCGAAATTCA AATTACTCGG CGTCCGCAGG ACTTTGACAA ACCACCGTAC
GGAATTAAGA GGGAGAGGAA TCCTAGTGGG AATAATTCAA GAACCGAGTT GGCTTTAAGT TTAATGAGCC GCAGGCGTCC TGAAACTGTT TGGTGGCATG

      2310       2320       2330       2340       2350       2360       2370       2380       2390       2400
TCCAAGTCTC GCTTCCGTCC CTACTTCAAA GTTATAACGT CCTCTTTATA AGTGATCTTG TTGAATTCTG GACAACCAAC CTTGGACTCC TACTAACCCT
AGGTTCAGAG CGAAGGCAGG GATGAAGTTT CAATATTGCA GGAGAAATAT TCACTAGAAC AACTTAAGAC CTGTTGGTTG GAACCTGAGG ATGATTGGGA

      2410       2420       2430       2440       2450       2460       2470       2480       2490       2500
CCACCGGTAA TCCTTAATAC GATTCTATAG AGACTCCGGA CTATTTGTTA TATCTTAATG ATTCTTCTAA TTATAATCTT TACGTACGAT GTATAGTCCC
GGTGGCCATT AGGAATTATG CTAAGATATC TCTGAGGCCT GATAAACAAT ATAGAATTAC TAAGAAGATT AATATTAGAA ATGCATGCTA CATATCAGGG

      2510       2520       2530       2540       2550       2560       2570       2580       2590       2600
TTACCCCGTC TCCAATATTA TCTATGTGTT CTATTTCGTC GAAAATCTAC AACATACTAC CCATACACCG GTCCCCAACA GCCGTACCTT CGTTATTGTG
AATGGGGCAG AGGTTATAAT AGATACACAA GATAAAGCAG CTTTTAGATG TTGTATGATG GGTATGTGGC CAGGGGTTGT CGGCATGGAA GCAATAACAC

      2610       2620       2630       2640       2650       2660       2670       2680       2690       2700
AATACTTATA ATCCAAATCT CCCCTACCCA TATTACCGTA ACATAAATAC CGATTGTGAT TCGACTAAGA TGTACCAACA TCGAAAAAAC CCAAATTATT
TTATGAATAT TAGGTTTAGA GGGGATGGGT ATAATGGCAT TGTATTTATG GCTAACACTA AGCTGATTCT ACATGGTTGT AGCTTTTTTG GGTTTAATAA

      2710       2720       2730       2740       2750       2760       2770       2780       2790       2800
ATGCACACAT CTTCGAACCC CCGTTCAATC ACACTCCCCA ACATCAAAAA TACGTACGAC CTAACGTTGT AGTCCATCCC ACTTCTCAGT CAACAGACAC
TACGTGTGTA GAAGCTTGGG GGCAAGTTAG TGTGAGGGGT TGTAGTTTTT ATGCATGCTG GATTGCAACA TCAGGTAGGG TGAAGAGTCA GTTGTCTGTG

      2810       2820       2830       2840       2850       2860       2870       2880       2890       2900
TTCTTTACGT ACAAACTCTC TACATTAGAA CCGTATGACT TACTTCCACT TCGTTCCCAG GCGGTGACGC GTCGATGTCT TTGACGGACG AAGTAAGATT
AAGAAATGCA TGTTTGAGAG ATGTAATCTT GGCATACTGA ATGAAGGTGA AGCAAGGGTC CGCCACTGCG CAGCTACAGA AACTGCCTGC TTCATTCTAA

      2910       2920       2930       2940       2950       2960       2970       2980       2990       3000
ATTTCCCTTT ACGGTCACAC TTCGTATTAT ACTAGACACC TGTAAGCCTA CTCTCCGGAA TAGTCTACGA TTGGACGCGA CCACCTGTAA CGTTATAAGA
C TAAAGGGAAA TGCCAGTGTG AAGCATAATA TGATCTGTGG ACATTCGGAT GAGAGGCCTT ATCAGATGCT AACCTGCGCT GGTGGACATT GCAATATTCT

      3010       3020       3030       3040       3050       3060       3070       3080       3090       3100
ACGATGGCAC GTATAGCAAA GTGTACGTGC GTTCTTTACC GGACATAAAC TTGTATTACA CTAATGGTTC ACGTGGTACG TATATCCACC AGCGTCCCCT
TGCTACCGTG CATATCGTTT CACATGCACG CAAGAAATGG CCTGTATTTG AACATAATGT GATTACCAAG TGCACCATGC ATATAGGTGG TCGCAGGGGA

      3110       3120       3130       3140       3150       3160       3170       3180       3190       3200
TACAAATACG GAATGGTCAC ATTGTACTTA GTACACTTCC ATTACAACCT TGGTCTACGG AAAAGGTCTC ACTCGCATTG TCCTTAGAAA CTATACTTAT
ATGTTTATGC CTTACCAGTG TAACATGAAT CATGTGAAGG TAATGTTGGA ACCAGATGCC TTTTCCAGAG TGAGCGTAAC AGGAATCTTT GATATGAATA

      3210       3220       3230       3240       3250       3260       3270       3280       3290       3300
AAGTTGATAC CTTCTAGGAC TCTATACTAC TGTGATTTGG TTCCCACGCG CGTACGCTTA CGCCTCCGTT CGTACGATCT AAGGTCGGCC ACACGCACCT
TTCAACTATG GAAGATCCTG AGATATGATG ACACTAAACC AAGGGTGCGC GCATGCGAAT GCGGAGGCAA GCATGCTAGA TTCCAGCCGG TGTGCGTGGA

      3310       3320       3330       3340       3350       3360       3370       3380       3390       3400
ACACTGACTT CTGGACTCCG GGCTAGTAAA CCACGAACGG ACGTGACCTC GCCTCAAGCC AAGATCACCA CTTCTTTGAC TGATTTCATT CATCACCCCC
TGTGACTGAA GACCTGAGGC CCGATCATTT GGTGCTTGCC TGCACTGGAG CGGAGTTCGG TTCTAGTGGT GAAGAAACTG ACTAAAGTAA GTAGTGGGGG
                                                                              stop 55 kD  ↑splice E1B RNA
                                                                              protein

      3410       3420       3430       3440       3450       3460       3470       3480       3490       3500
GTTTTACACC TACCCCTGAA AGTCCAACCA TTCCACCTGT TTAACCCATT TAAAACAATT AAAGACAGAA CGTCGACGGT ACTCACCTTC GCGAAGAAAA
CAAAATGTGG ATGGGGACTT TCAGGTTGGT AAGGTGGACA AATTGGGTAA ATTTTGTTAA. TTTCTGTCTT GCAGCTGCCA TGAGTGGAAG CGCTTCTTTT
                                                        start ↑protein IX RNA splice ↑ start   stop 9 kD protein
                                                                              E1B RNA protein IX

      3510       3520       3530       3540       3550       3560       3570       3580       3590       3600
CTCCCCCCTC ATAAATCGGG AATAGACTGC CCGTCCGAGG GTGGTACCCG TCCTCAAGCA GTCTTACAGT ACCCTAGGTG ACACCTACCC TCTGGGCAGG
GAGGGGGGAG TATTTAGCCC TTATCTGACG GGCAGGCTCC CACCATGGGC AGGAGTTCGT CAGAATGTCA TGGGATCCAC TGTGGATGGG AGACCCGTCC

      3610       3620       3630       3640       3650       3660       3670       3680       3690       3700
TCGGGCGGTT AAGGAGTTGC GACTGGATAC GGTGAAACTC AAGCAGTGGT AACCTACGTC GACGTCGGCG GCGGCGATGA CGACGGCGGT TGTGGTAGGA
AGCCCGCCAA TTCCTCAACG CTGACCTATG CCACTTTGAG TTCGTCACCA TTGGATGCAG CTGCAGCCGC CGCCGCTACT GCTGCCGCCA ACACCATCCT

      3710       3720       3730       3740       3750       3760       3770       3780       3790       3800
ACCTTACCCG ATAATGCCTT CGTAACAACG GTTAAGGTCA AGGAGATTAT TAGGAAGTTG GGACCGACTC CTGTTCGATG AACAAGAGAA CCGAGTCGAG
TGGAATGGGC TATTACGGAA GCATTGTTGC CAATTCCAGT TCCTCTAATA ATCCTTCAAC CCTGGCTGAG GACAAGCTAC TTGTTCTCTT GGCTCAGCTC

      3810       3820       3830       3840       3850       3860       3870       3880       3890       3900
CTCCGGAATT GGGTTGCGAA TCCGCTTGAC AGATTCGTCC ACCGGGTCAA CGCACTCGTT TGACTCAGAC GACAACGGTG TCGTTTCAGA TTTATTTCTA
GAGGCCTTAA CCCAACGCTT AGGCGAACTG TCTAAGCAGG TGGCCCAGTT GCGTGAGCAA ACTGAGTCTG CTGTTGCCAC AGCAAAGTCT AAATAAAGAT
                                                                                                stop protein IX

                                                 stop protein IVa2
      3910       3920       3930       3940       3950       3960       3970       3980       3990       4000
GAGTTTAGTT ATTTATTTCT TTATGAACAA TATTTTTGTT TACTTACAAA TAAACTAAAA AGCGCGCGCC ATACGGGACC TGGTAGCCAA AGCTAGTAAC
D CTCAAATCAA TAAATAAAGA AATACTTGTT ATAAAAACAA ATGAATGTTT ATTTGATTTT TCGCGCGCGG TATGCCCTGG ACCATCGGTT TCGATCATTG
                                  poly A ↑E1B RNAs
```

FIGURE 26.1 (*Continued*)

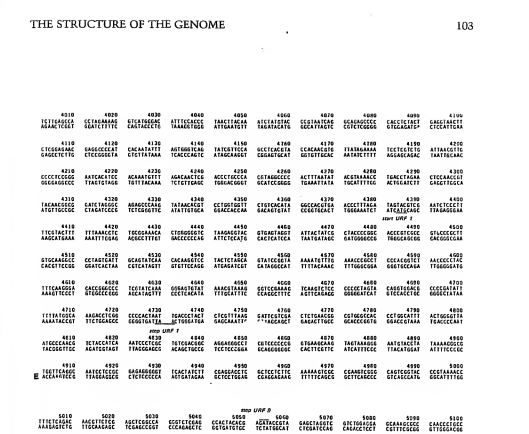

FIGURE 26.1 (*Continued*)

FIGURE 26.1 (*Continued*)

FIGURE 26.1 (*Continued*)

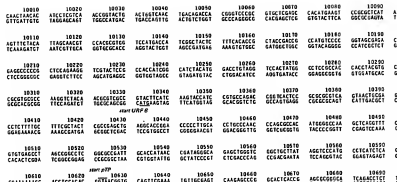

FIGURE 26.1 (*Continued*)

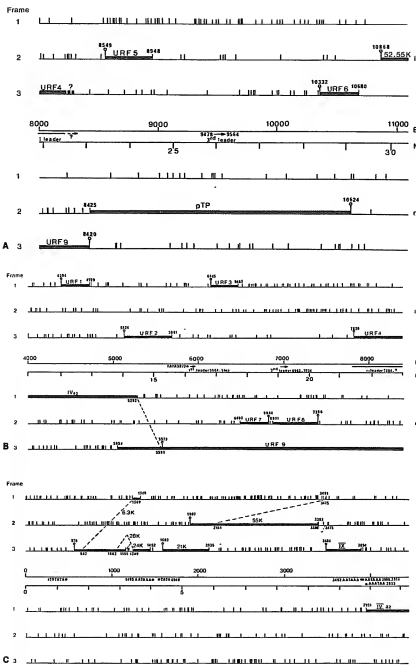

FIGURE 26.2A–C. Structural organization of a region between coordinates 0 and 31.7 on the Ad7 genome. This map is derived from the nucleotide sequence in Fig. 26.1. For details, see Fig. 3 [Section VII].

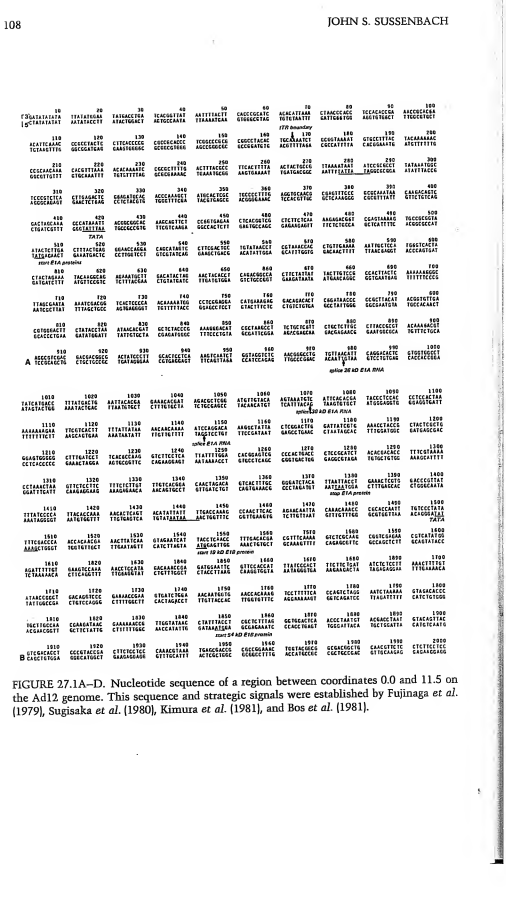

FIGURE 27.1A–D. Nucleotide sequence of a region between coordinates 0.0 and 11.5 on the Ad12 genome. This sequence and strategic signals were established by Fujinaga *et al.* (1979), Sugisaka *et al.* (1980), Kimura *et al.* (1981), and Bos *et al.* (1981).

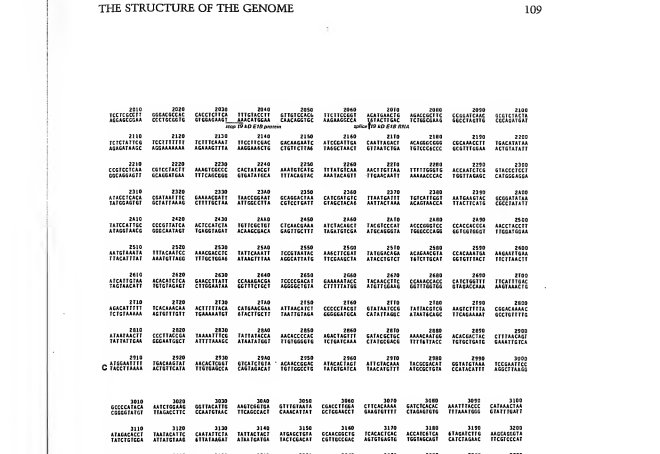

FIGURE 27.1 (*Continued*)

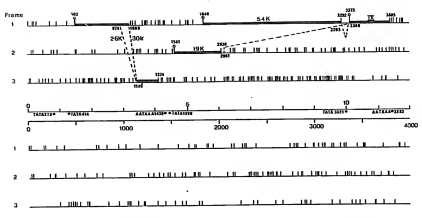

FIGURE 27.2. Structural organization of a region between coordinates 0.0 and 11.5 on the Ad12 genome. This map is derived from the nucleotide sequence in Fig. 27.1.

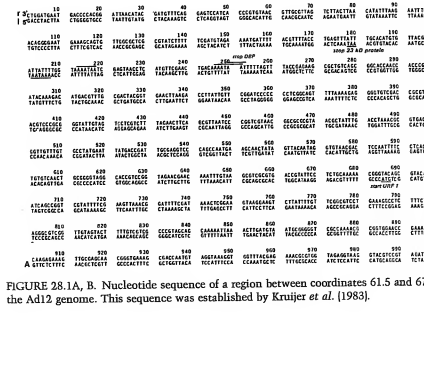

FIGURE 28.1A, B. Nucleotide sequence of a region between coordinates 61.5 and 67.0 on the Ad12 genome. This sequence was established by Kruijer *et al.* (1983).

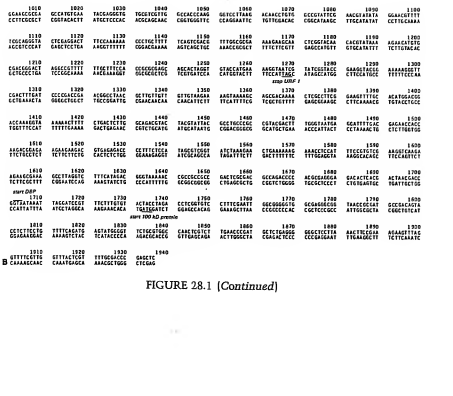

FIGURE 28.1 (*Continued*)

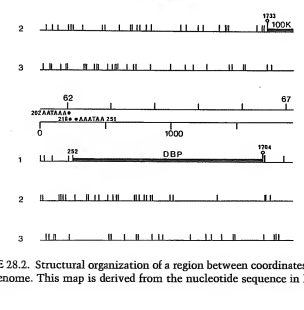

FIGURE 28.2. Structural organization of a region between coordinates 61.5 and 67.0 on the Ad12 genome. This map is derived from the nucleotide sequence in Fig. 28.1.

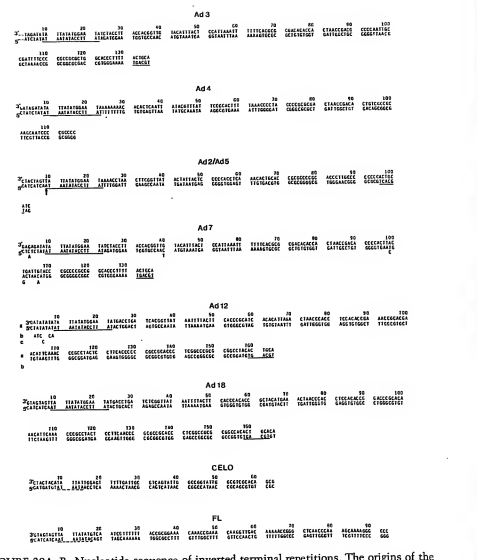

FIGURE 29A, B. Nucleotide sequence of inverted terminal repetitions. The origins of the sequences are as follows: (A) Ad3: Tolun *et al.* (1979). Ad4: Tokunaga *et al.* (1982). Ad2/Ad5: The Ad2 sequence was determined by Shinagawa and Padmanabhan (1979) and the Ad5 sequence by Steenbergh *et al.* (1977). The two sequences are identical. Arrand and Roberts (1979) have analyzed an Ad2 strain that missed base pair 9 (↑). Ad7: These sequences were determined for strain Gomen by Dijkema and Dekker (1979) [a] and for strain Greider by Shinagawa and Padmanabhan (1980) [b]. The differences between the sequences are indicated. (B) Ad12: Tolun *et al.* (1979) [a], Sugisaka *et al.* (1980) [a], Shinagawa and Padmanabhan (1980) [b], and Schwarz *et al.* (1982) [c]. The differences between the sequences are indicated. Ad18: Garon *et al.* (1982). CELO: Alestrőm *et al.* (1982a). FL: Temple *et al.* (1981). In the human sequences, the conserved sequences 9–22 are underlined; the homologous regions in CELO and FL DNA are indicated by dashed underlines. The common sequence TGACGT discovered by Shinagawa and Padmanabhan (1980) is underlined.

REFERENCES

Akusjärvi, G., and Persson, H., 1981a, Gene and mRNA for precursor polypeptide VI from adenovirus type 2, *J. Virol.* **38**:469.

Akusjärvi, G., and Persson, H., 1981b, Control of RNA splicing and termination in the major late adenovirus transcription unit, *Nature [London]* **292**:420.

Akusjärvi, G., and Pettersson, U., 1978a, Sequence analysis of adenovirus DNA. I. Nucleotide sequence at the carboxy-terminal end of the gene for adenovirus type 2 hexon, *Virology* **91**:477.

Akusjärvi, G., and Pettersson, U., 1978b, Nucleotide sequence at the junction between the coding region of the adenovirus 2 hexon messenger RNA and its leader sequence, *Proc. Natl. Acad. Sci. U.S.A.* **75**:5822.

Akusjärvi, G., and Pettersson, U., 1979a, Sequence analysis of adenovirus DNA: Complete nucleotide sequence of the spliced 5'-non-coding region of adenovirus 2 hexon messenger RNA, *Cell* **16**:841.

Akusjärvi, G., and Pettersson, U., 1979b, Sequence analysis of adenovirus DNA. IV. The genomic sequences encoding the common tripartite leader of late adenovirus messenger RNA, *J. Mol. Biol.* **13**:143.

Akusjärvi, G., Mathews, M.B., Anderson, P., Vennström, B., and Pettersson, U., 1980, Structure of genes for virus-associated RNAI and RNAII of adenovirus type 2, *Proc. Natl. Acad. Sci. U.S.A.* **77**:2424.

Akusjärvi, G., Zabielsky, J., Perricaudet, M., and Pettersson, U., 1981, The sequence of the 3' non-coding region of the hexon mRNA discloses a novel adenovirus gene, *Nucleic Acids Res.* **9**:1.

Aleström, P., Akusjärvi, G., Perricaudet, M., Mathews, M.B., Klessig, D., and Pettersson, U., 1980, Sequence analysis of adenovirus DNA. VI. The nucleotide sequence of the gene for polypeptide IX from adenovirus type 2 and its colinear messenger RNA, *Cell* **19**:671.

Aleström, P., Stenlund, A., Li, P., and Pettersson, U., 1982a, A common sequence in the inverted terminal repetitions of human and avian adenoviruses, *Gene* **18**:193.

Aleström, P., Akusjärvi, G., Pettersson, M., and Pettersson, U., 1982b, DNA sequence analysis of the region encoding the terminal protein and the hypothetical N-gene product of adenovirus type 2, *J. Biol. Chem.* **257**:13492.

Anderson, C.W., and Lewis, J.B., 1980, Amino-terminal sequence of the adenovirus type 2 proteins: Hexon, fiber, component IX and the early protein IB-15 K, *Virology* **104**:27.

Ariga, H., Shimojo, H., Hidaka, S., and Miura, M., 1979, Specific cleavage of the terminal protein from the adenovirus 5 DNA under the conditions of single-strand scission by nuclease S1, *FEBS Lett.* **107**:355.

Ariga, H., Klein, H., Levine, A.J., and Horwitz, M.S., 1980, A cleavage product of the adenovirus DNA binding protein is active for DNA replication *in vitro*, *Virology* **101**:307.

Arrand, J.R., and Roberts, R.J., 1979, The nucleotide sequences at the termini of adenovirus 2 DNA, *J. Mol. Biol.* **128**:577.

Arrand, J.R., Keller, W., and Roberts, R.J., 1975, Extent of terminal repetition in adenovirus 2 DNA, *Cold Spring Harbor Symp. Quant. Biol.* **39**:401.

Baker, C.C., and Ziff, E.B., 1980, Biogenesis, structures, and sites of encoding of the 5' termini of adenovirus-2 mRNAs, *Cold Spring Harbor Symp. Quant. Biol.* **44**:415.

Baker, C.C., and Ziff, E.B., 1981, Promoters and heterogeneous 5'-termini of the messenger RNAs of adenovirus-2, *J. Mol. Biol.* **149**:189.

Baker, C.C., Hérissé, J., Courtois, G., Galibert, F., and Ziff, E., 1979, Messenger RNA for the Ad2 DNA binding protein: DNA sequences encoding the first leader and heterogeneity at the mRNA 5' end, *Cell* **8**:569.

Bartok, K., Garon, C.F., Berry, K.W., and Rose, J.A., 1974, Specific fragmentation of adenovirus heteroduplex DNA molecules with single-strand specific nuclease of *Neurospora crassa*, *J. Mol. Biol.* **87**:437.

Bégin, M., and Weber, J., 1975, Genetic analysis of adenovirus type 2. I. Isolation and genetic characterization of temperature-sensitive mutants, *J. Virol.* **115**:1.

Berget, S.M., Moore, C., and Sharp, P.A., 1977, Spliced segments at the 5′ terminus of adenovirus 2 late mRNA, *Proc. Natl. Acad. Sci. U.S.A.* **74**:3171.

Berget, S.M., Berk, A., Harrison, T., and Sharp, P.A., 1978, Spliced segments at the 5′ termini of adenovirus 2 late mRNA: A role for heterogeneous nuclear RNA in mammalian cells, *Cold Spring Harbor Symp. Quant. Biol.* **42**:523.

Berk, A.J., and Sharp, P.A., 1977a, Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease digested hybrids, *Cell* **12**:721.

Berk, A.J., and Sharp, P.A., 1977b, UV mapping of the adenovirus 2 early promoters, *Cell* **12**:45.

Berk, A.J., and Sharp, P.A., 1978, Structure of the adenovirus 2 early mRNAs, *Cell* **14**:695.

Berk, A.J., Lee, F., Harrison, T., Williams, J., and Sharp, P.A., 1979, Pre-early adenovirus 5 gene product regulates synthesis of early viral messenger RNA, *Cell* **17**:935.

Bernards, R., Houweling, A., Schrier, P.I., Bos, J.L., and van der Eb, A.J., 1982, Characterization of cells transformed by Ad5/Ad12 hybrid early region 1 plasmids, *Virology* **120**:422.

Bhatti, A.R., and Weber, J., 1979, Protease of adenovirus type 2, *J. Biol. Chem.* **254**:12265.

Binger, M.H., Flint, S.J., and Rekosh, D.M., 1982, Expression of the gene encoding the adenovirus DNA terminal protein precursor in productively infected and transformed cells, *J. Virol.* **42**:488.

Bos, J.L., Polder, L.J., Bernards, R., Schrier, P.I., van den Elsen, P., Van der Eb, A.J., and van Ormondt, H., 1981, The 2.2 kb mRNA of human Ad12 and Ad5 codes for two tumor antigens starting at different AUG triplets, *Cell* **27**:121.

Brackman, K.H., Green, M., Wold, W.S.M., Cartas, M., Matsuo, T., and Hashimoto, S., 1980, Identification and peptide mapping of human adenovirus type 2-induced early polypeptides isolated by two-dimensional gel electrophoresis, *J. Biol. Chem.* **255**:6772.

Broker, T.R., Chow, L.T., Dunn, A.R., Gelinas, R.E., Hassell, J.A., Klessig, D.F., Lewis, J.B., Roberts, R.J., and Zain, B.S., 1977, Adenovirus-2 messengers—An example of baroque molecular architecture, *Cold Spring Harbor Symp. Quant. Biol.* **42**:531.

Brown, D.T., Westphal, M., Burlingham, B.T., Winterhoff, U., and Doerfler, W., 1975, Structure and composition of the adenovirus type 2 core, *J. Virol.* **16**:366.

Burnett, J.P., and Harrington, J.A., 1968, Simian adenovirus SA7 DNA: Chemical, physical and biological studies, *Proc. Natl. Acad. Sci. U.S.A.* **60**:1023.

Carter, T.H., and Blanton, R.A., 1978, Possible role of the 72,000-dalton DNA-binding protein in regulation of adenovirus type 5 early gene expression, *J. Virol.* **25**:664.

Carter, T.H., and Ginsberg, H.S., 1976, Viral transcription in KB cells infected by temperature-sensitive early mutants of adenovirus type 5, *J. Virol.* **18**:156.

Carusi, E.A., 1977, Evidence for blocked 5′ termini in human adenovirus DNA, *Virology* **76**:380.

Celma, M.L., Pan, J., and Weissmann, S., 1977a, Studies of low molecular weight RNA from cells infected with adenovirus 2. I. The sequence at the 3′ end of VA-RNA 1, *J. Biol. Chem.* **252**:9032.

Celma, M.L., Pan, J., and Weissmann, S., 1977b, Studies of low molecular weight RNA from cells infected with adenovirus 2. II. Heterogeneity at the 5′ end of VA-RNA 1, *J. Biol. Chem.* **252**:9043.

Challberg, M.D., and Kelly, T.J., Jr., 1979a, Adenovirus DNA replication *in vitro*, *Proc. Natl. Acad. Sci. U.S.A.* **76**:655.

Challberg, M.D., and Kelly, T.J., Jr., 1979b, Adenovirus DNA replication *in vitro*: Origin and direction of daughter strand synthesis, *J. Mol. Biol.* **135**:999.

Challberg, M.D., and Kelly, T.J., Jr., 1981, Processing of the adenovirus terminal protein, *J. Virol.* **38**:272.

Challberg, M.D., Desiderio, S.V., and Kelly, T.J., Jr., 1980, Adenovirus DNA replication *in vitro*: Characterization of a protein covalently linked to nascent DNA strands, *Proc. Natl. Acad. Sci. U.S.A.* **77**:5105.

Chinnadurai, G., Chinnadurai, S., and Green, M., 1978, Enhanced infectivity of adenovirus DNA and a DNA–protein complex, *J. Virol.* 26:195.

Chow, L.T., and Broker, T.R., 1978, The spliced structures of adenovirus-2 fiber message and the other mRNAs, *Cell* 15:497.

Chow, L.T., Gelinas, R.E., Broker, T.R., and Roberts, R.J., 1977a, An amazing sequence arrangement at the 5′ ends of adenovirus 2 messenger RNA, *Cell* 12:1.

Chow, L.T., Roberts, J.M., Lewis, J.B., and Broker, T.R., 1977b, A cytoplasmic RNA transcript map from adenovirus type 2, determined by electron microscopy of RNA:DNA hybrids, *Cell* 11:819.

Chow, L.T., Broker, T.R., and Lewis, J.B., 1979a, Complex splicing patterns of RNAs from the adenovirus 2, *J. Mol. Biol.* 134:265.

Chow, L., Lewis, J.B., and Broker, T., 1979b, RNA transcription and splicing at early and intermediate times after adenovirus-2 infection, *Cold Spring Harbor Symp. Quant. Biol.* 44:401.

Coombs, D.H., and Pearson, G.D., 1978, Filter-binding assay for covalent DNA–protein complexes: Adenovirus DNA–terminal protein complex, *Proc. Natl. Acad. Sci. U.S.A.* 75:5291.

Coombs, D.H., Robinson, A.J., Bodnar, J.W., Jones, C.J., and Pearson, G.D., 1978, Detection of DNA–protein complexes: The adenovirus DNA–terminal protein and HeLa DNA–protein complexes, *Cold Spring Harbor Symp. Quant. Biol.* 43:741.

Corden, J., Engelking, H.M., and Pearson, G.D., 1976, Chromatin-like organization of the adenovirus chromosome, *Proc. Natl. Acad. Sci. U.S.A.* 73:401.

Desiderio, S.V., and Kelly, T., 1981, Structure of the linkage between adenovirus DNA and the 55,000 molecular weight terminal protein, *J. Mol. Biol.* 145:319.

Dijkema, R., and Dekker, B.M.M., 1979, The inverted terminal repetition of the DNA of the weakly oncogenic adenovirus type 7, *Gene* 8:7.

Dijkema, R., Dekker, B.M.M., and van Ormondt, H., 1980a, The nucleotide sequence of the transforming BglII-H fragment of adenovirus type 7 DNA, *Gene* 9:141.

Dijkema, R., Dekker, B.M.M., van Ormondt, H., Maat, J., and Boyer, H., 1980b, Gene organization of the transforming region of weakly oncogenic adenovirus type 7: The E1a region, *Gene* 12:287.

Dijkema, R., Maat, J., Dekker, B.M.M., van Ormondt, H., and Boyer, W., 1981, The gene for polypeptide IX of adenovirus type 7, *Gene* 13:375.

Dijkema, R., Dekker, B.M.M., and van Ormondt, H., 1982, Gene organization of the transforming region of adenovirus type 7, *Gene* 18:143.

Dimitrov, D.H., Dubitchev, A.G., Naroditsky, B.S., Dreizin, R.S., and Tikchonenko, T.I., 1979, Physicochemical properties and restriction maps of simian adenovirus 38, *J. Gen. Virol.* 44:69.

Doerfler, W., and Kleinschmidt, A., 1970, Denaturation pattern of the DNA of adenovirus type 2 as determined by electron microscopy, *J. Mol. Biol.* 50:579.

Doerfler, W., Hellmann, W., and Kleinschmidt, A.K., 1972, The DNA of adenovirus type 12 and its denaturation pattern, *Virology* 47:507.

Dunn, A.R., Mathews, M.B., Chow, L.T., Sambrook, J., and Keller, W., 1978, A supplementary adenoviral leader sequence and its role in messenger translation, *Cell* 15:511.

Ellens, D.J., Sussenbach, J.S., and Jansz, H.S., 1974, Studies on the mechanism of replication of adenovirus DNA. III. Electron microscopy of replicating DNA, *Virology* 61:427.

Engler, J., 1981, The nucleotide sequence of the polypeptide IX gene of human adenovirus type 3, *Gene* 13:387.

Engler, J.A., and van Bree, M., 1982, The nucleotide sequence of the gene encoding protein IVa2 in human adenovirus type 7, *Gene* 19:71.

Engler, J., Broker, T.R., and Chow, L.T., 1981, Sequences of human adenovirus Ad3 and Ad7 DNAs encoding the promoter and first leader segment of late RNAs, *Gene* 13:133.

Engler, J.A., Hoppe, M.S., and van Bree, M.P., 1983, The nucleotide sequence of the genes encoded in the early region 2B in human adenovirus type 7, *Gene* 21:147.

Enomoto, T., Lichy, J.H., Ikeda, J.E., and Hurwitz, J., 1981, Adenovirus DNA replication *in*

vitro: Purification of the terminal protein in a functional form, *Proc. Natl. Acad. Sci. U.S.A.* 78:6779.

Esche, H., Mathews, M.B., and Lewis, J.B., 1980, Proteins and messenger RNAs of the transforming region of wild-type and mutant adenoviruses, *J. Mol. Biol.* 142:399.

Evans, R.M., Fraser, N., Ziff, E., Weher, J., Wilson, M., and Darnell, J.E., 1977, The initiation sites for RNA transcription in Ad2 DNA, *Cell* 12:733.

Everitt, E., Sundquist, B., Petersson, O., and Philipson, L., 1973, Structural proteins of adenoviruses. X. Isolation and topography of low molecular weight antigens from the virion of adenovirus type 2, *Virology* 62:130.

Forsblom, S., Rigler, R., Ehrenberg, M., Pettersson, U., and Philipson, L., 1976, Kinetic studies on the cleavage of adenovirus DNA by restriction endonuclease EcoRI, *Nucleic Acids Res.* 3:3255.

Fraser, N., and Ziff, E., 1978, RNA structures near poly (A) of adenovirus-2 late messenger RNA's, *J. Mol. Biol.* 124:27.

Fraser, N.W., Nevins, J.R., Ziff, E., and Darnell, J.E., 1979, The major late adenovirus type 2 transcription unit: Termination is downstream from the last poly (A) site, *J. Mol. Biol.* 129:643.

Fraser, N.W., Baker, C.C., Moore, M.A., and Ziff, E.B., 1982, Poly(A) sites of adenovirus serotype 2 transcription units, *J. Mol. Biol.* 155:207.

Freeman, A.E., Black, P.E., Vanderpool, E.A., Henry, P.H., Austin, J.B., and Huebner, R.J., 1967, Transformation of primary rat embryo cells by adenovirus type 2, *Proc. Natl. Acad. Sci. U.S.A.* 58:1205.

Frost, E., and Williams, J., 1978, Mapping of temperature-sensitive and host-range mutants of adenovirus type 5 by marker rescue, *Virology* 91:39.

Fujinaga, K., Sawada, Y., Uemizu, Y., Yamashita, T., Shimojo, H., Shiroki, K., Sugisaki, H., Sugimoto, K., and Takanami, M., 1979, Nucleotide sequences, integration, and transcription of the adenovirus-12 transforming genes, *Cold Spring Harbor Symp. Quant. Biol.* 44:519.

Galibert, F., Hérissé, J., and Courtois, G., 1979, Nucleotide sequence of the EcoRI-F fragment of adenovirus 2 genome, *Gene* 6:1.

Gallimore, P.H., 1974, Interactions of adenovirus type 2 with rat embryo cells: Permissiveness, transformation and *in vitro* characteristics of adenovirus transformed rat embryo cells, *J. Gen. Virol.* 25:263.

Galos, R.S., Williams, J., Binger, M.H., and Flint, S.J., 1979, Location of additional early gene sequences in the adenoviral chromosome, *Cell* 17:945.

Gambke, C., and Deppert, W., 1981, Late non-structural 100,000 and 33,000 dalton proteins of adenovirus type 2. II. Immunological and protein chemical analysis, *J. Virol.* 40:594.

Garon, C.F., Berry, K.W., and Rose, J.A., 1972, A unique form of terminal redundancy in adenovirus DNA molecules, *Proc. Natl. Acad. Sci. U.S.A.* 69:2391.

Garon, C.F., Berry, K.W., Hierholzer, J.C., and Rose, J.A., 1973, Mapping base sequence heterologies between genomes from different adenovirus serotypes, *Virology* 54:414.

Garon, C.F., Berry, K.W., and Rose, J.A., 1975, Arrangement of sequences in the inverted terminal repetition of adenovirus 18 DNA, *Proc. Natl. Acad. Sci. U.S.A.* 72:3039.

Garon, F.C., Parr, R.P., Padmanabhan, R., Roninson, I., Garrison, J.W., and Rose, J.A., 1982, Structural characterization of the adenovirus 18 inverted terminal repetition, *Virology* 121:230.

Gingeras, T.R., Sciaky, D., Gelinas, R.E., Bing-Dong, J., Yen, C.E., Kelly, M.M., Bullock, P.A., Parsons, B.L., O'Neill, K.E., and Roherts, R.J., 1982, Nucleotide sequences from the adenovirus-2 genome, *J. Biol. Chem.* 257:13475.

Ginsberg, H.S., Ensinger, M.J., Kauffman, R.S., Mayer, A.J., and Lundholm, U., 1974, Cell transformation: A study of regulation with types 5 and 12 adenovirus temperature-sensitive mutants, *Cold Spring Harbor Symp. Quant. Biol.* 44:419.

Girardi, A.J., Hilleman, M.R., and Zwickey, R.E., 1964, Tests in hamsters for oncogenic quality of ordinary viruses including adenovirus type 7, *Proc. Soc. Exp. Biol. Med.* 115:1141.

Goldbach, R.W., Evers, R.F., and Borst, P., 1978, Electrophoretic strand separation of long DNAs with poly-UG in agarose gels, *Nucleic Acids Res.* **5**:2743.

Goodhearst, C.R., 1971, DNA density of oncogenic and non-oncogenic simian adenoviruses, *Virology* **44**:645.

Graham, F.L., and van der Eb, A.J., 1973, A new technique for the assay of the infectivity of human adenovirus 5 DNA, *Virology* **52**:456.

Graham, F.L., van der Eb, A.J., and Heijneker, H.L., 1974a, Size and location of the transforming regions of human adenovirus type 5 DNA, *Nature (London)* **251**:687.

Graham, F.L., Abrahams, P.J., Mulder, C., Heijneker, H.L., Warnaar, S.O., de Vries, F.A.J., Fiers, W., and van der Eb, A.J., 1974b, Studies on *in vitro* transformation by DNA and DNA fragments of human adenoviruses and simian virus 40, *Cold Spring Harbor Symp. Quant. Biol.* **39**:637.

Green, M., 1970, Oncogenic viruses, *Annu. Rev. Biochem.* **39**:701.

Green, M., and Piña, M., 1963, Biochemical studies on adenovirus multiplication. IV. Isolation, purification and chemical analysis of adenovirus, *Virology* **20**:199.

Green, M., and Piña, M., 1964, Biochemical studies on adenovirus multiplication. VI. Properties of highly purified tumorigenic human adenoviruses and their DNA's, *Proc. Natl. Acad. Sci. U.S.A.* **51**:1251.

Green, M., Piña, M., Kimes, R., Wensink, P.C., Machattie, L.A., and Thomas, C.A., 1967, Adenovirus DNA. I. Molecular weight and conformation, *Proc. Natl. Acad. Sci. U.S.A.* **57**:1302.

Green, M., Parsons, J.T., Piña, M., Fujinaga, K., Caffier, H., and Landgraf-Leurs, I., 1970, Transcription of adenovirus genes in productively infected and transformed cells, *Cold Spring Harbor Symp. Quant. Biol.* **30**:803.

Green, M., Wold, W.S.M., Brackmann, K.H., and Cartas, M.A., 1979a, Identification of families of overlapping polypeptides coded by early "transforming" gene region 1 of human adenovirus type 2, *Virology* **97**:275.

Green, M., Mackey, J., Wold, W.S.M., and Rigden, P., 1979b, Thirty-one human adenovirus serotypes (Ad1–Ad31) form five groups (A–E) based on upon DNA genome homologies, *Virology* **93**:481.

Green, M., Brackmann, K., Wold, W.S.M., Cartas, M., Thornton, H., and Elder, J.H., 1979c, Conserved primary sequences of the DNA terminal proteins of five different human adenovirus groups, *Proc. Natl. Acad. Sci. U.S.A.* **76**:4380.

Green, M., Wold, W.S.M., Brackmann, K., and Cartas, M.A., 1979d, Studies on early proteins and transformation proteins in human adenoviruses, *Cold Spring Harbor Symp. Quant. Biol.* **44**:457.

Grodzicker, T., Williams, J., Sharp, P.A., and Sambrook, J., 1975, Physical mapping of temperature-sensitive mutations of adenoviruses, *Cold Spring Harbor Symp. Quant. Biol.* **39**:439.

Grodzicker, T., Anderson, C.W., Sambrook, J., and Mathews, M.B., 1977, The physical location of structural genes in adenovirus DNA, *Virology* **80**:111.

Harrison, T., Graham, F.L., and Williams, J., 1977, Host range mutants of adenovirus type 5 defective for growth in HeLa cells, *Virology* **77**:319.

Harter, M.L., and Lewis, J.B., 1978, Adenovirus type 2 early proteins synthesized *in vitro* and *in vivo*: Identification in infected cells of the 38,000- to 50,000-molecular-weight protein encoded by the left end of adenovirus type 2 genome, *J. Virol.* **26**:736.

Harter, M.L., Shanmugan, G., Wold, W.S.M., and Green, M., 1976, Detection of adenovirus type 2 induced early polypeptides using cycloheximide pretreatment to enhance viral protein synthesis, *J. Virol.* **19**:1976.

Hashimoto, S., Pursley, M.H., and Green, M., 1981, Nucleotide sequences and mapping of novel heterogeneous 5'-termini of adenovirus type 2 early region 4 mRNA, *Nucleic Acids Res.* **9**:1675.

Hérissé, J., and Galibert, F., 1981, Nucleotide sequence of the EcoRI E fragment of adenovirus-2 genome, *Nucleic Acids Res.* **9**:1229.

Hérissé, J., Courtois, G., and Galibert, F., 1980, Nucleotide sequence of the EcoRI D fragment of adenovirus 2 genome, *Nucleic Acids Res.* **8**:2173.

Hérissé, J., Rigolet, M., Dupont de Dinechin, S., and Galibert, F., 1981, Nucleotide sequence of adenovirus 2 fragment encoding for the carboxylic region of the fiber protein and the entire E4 region, *Nucleic Acids Res.* 9:4023.

Hierholzer, J.C., 1973, Further subgrouping of the human adenoviruses by differential hemagglutination, *J. Infect. Dis.* 128:541.

Horwitz, S., 1974, Location of the origin of DNA replication in adenovirus type 2, *J. Virol.* 13:1046.

Horwitz, M.S., 1978, Temperature-sensitive replication of H5ts125 adenovirus DNA in the presence of cycloheximide, *J. Virol.* 11:544.

Houweling, A., van den Elsen, P.J., and van der Eb, A.J., 1980, Partial transformation of primary rat cells by the leftmost 4.5% fragment of adenovirus 5 DNA, *Virology* 105:537.

Huebner, R.J., Rowe, W.P., and Lane, W.T., 1962, Oncogenic effects in hamsters of human adenovirus types 12 and 18, *Proc. Natl. Acad. Sci. U.S.A.* 48:2051.

Huebner, R.J., Casey, M.J., Chanock, R.M., and Schell, K., 1965, Tumors induced in hamsters by a strain of adenovirus type 3: Sharing of tumor antigens and neoantigens with those produced by adenovirus type 7 tumors, *Proc. Soc. Exp. Biol. Med.* 54:381.

Jones, N., and Shenk, T., 1979a, An adenovirus type 5 early gene function regulates expression of other early viral genes, *Proc. Natl. Acad. Sci. U.S.A.* 76:3665.

Jones, N., and Shenk, T., 1979b, Isolation of adenovirus type 5 host range deletion mutants defective for transformation of rat embryo cells, *Cell* 17:683.

Jörnvall, H., Akusjärvi, G., Aleström, P., von Bahr-Lindstrom, H., Pettersson, U., Appella, E., Fowler, A.V., and Philipson, L., 1981a, The adenovirus hexon protein: The primary structure of the polypeptide and its correlation with the hexon gene, *J. Biol. Chem.* 256:6181.

Jörnvall, H., Aleström, P., Akusjärvi, G., von Bahr-Lindstrom, H., Philipson, L., and Pettersson, U., 1981b, Order of the CNBr fragments in the adenovirus hexon protein, *J. Biol. Chem.* 256:6204.

Keegstra, W.S., van Wielink, P.S., and Sussenbach, J.S., 1977, The visualization of a circular DNA–protein complex from adenovirions, *Virology* 76:444.

Kilpatrick, B.A., Gelinas, R.E., Broker, T.R., and Chow, L.T., 1979, Comparison of late mRNA splicing among class B and class C adenoviruses, *J. Virol.* 30:899.

Kimes, R., and Green, M., 1970, Adenovirus DNA. II. Separation of molecular halves of adenovirus type 2, *J. Mol. Biol.* 50:203.

Kimura, T., Sawada, Y., Shinagawa, M., Shimizu, Y., Shiroki, K., Shimojo, H., Sugisaki, H., Takanami, M., Uemizu, Y., and Fujinaga, K., 1981, Nucleotide sequence of the transforming early region E1b of adenovirus 12 DNA: Structure and gene organization and comparison with those of adenovirus type 5 DNA, *Nucleic Acids Res.* 9:6571.

Kitchingman, G.R., and Westphal, H., 1980, The structure of adenovirus 2 early nuclear and cytoplasmics RNAs. *J. Mol. Biol.* 137:23.

Kitchingman, G.R., Lai, S.-P., and Westphal, H., 1977, Loop structures in hybrids of early RNA and the separated strands of adenovirus DNA, *Proc. Natl. Acad. Sci. U.S.A.* 74:4329.

Klessig, D.F., and Grodzicker, T., 1979, Mutations that allow human Ad2 and Ad5 to express late genes in monkey cells map in the viral gene encoding the 72K DNA binding protein, *Cell* 17:957.

Kruijer, W., van Schaik, F.M.A., and Sussenbach, J.S., 1980, Nucleotide sequence analysis of a region of adenovirus 5 DNA encoding a hitherto unidentified gene, *Nucleic Acids Res.* 8:6033.

Kruijer, W., van Schaik, F.A.M., and Sussenbach, J.S., 1981, Structure and organization of the gene coding for the DNA binding protein of adenovirus type 5, *Nucleic Acids Res.* 9:4439.

Kruijer, W., van Schaik, F.A.M., and Sussenbach, J.S., 1982, Nucleotide sequence of the gene encoding adenovirus type 2 DNA binding protein, *Nucleic Acids Res.* 10:4493.

Kruijer, W., van Schaik, F.A.M., Speyer, J.G., and Sussenbach, J.S., 1983, Structure and function of denovirus DNA binding protein: comparison of the amino acid sequences of

Ad5 and Ad12 proteins derived from the nucleotide sequence of the corresponding genes, *Virology* **128**:140.

Kubinski, H., and Rose, J.A., 1967, Regions containing repeating base-pairs in DNA from some oncogenic and non-oncogenic animal viruses, *Proc. Natl. Acad. Sci. U.S.A.* **57**:1720.

Landgraf-Leurs, M., and Green, M., 1971, Adenovirus DNA. III. Separation of the complementary strands of adenovirus types 2, 7 and 12 DNA molecules, *J. Mol. Biol.* **60**:185.

Larsen, S.H., Margolskee, R.F., and Nathans, D., 1979, Alignment of the restriction map of mouse adenovirus FL and human adenovirus type 2, *Virology* **97**:406.

Larson, V.M., Girardi, A.J., Hilleman, M.R., and Zwickey, R.E., 1965, Studies on oncogenicity of adenovirus type 7 viruses in hamsters, *Proc. Soc. Exp. Biol. Med.* **118**:15.

Laver, W.G., 1970, Isolation of an arginine-rich protein from particles of adenovirus type 2, *Virology* **41**:488.

Laver, W.G., Suriano, J.R., and Green, M., 1967, Adenovirus proteins. II. N-terminal amino acid analysis, *J. Virol.* **1**:723.

Laver, W.G., Pereira, H.G., Russell, W.C., and Valentine, R.C., 1968, Isolation of an internal component of adenovirus type 5, *J. Mol. Biol.* **37**:379.

Laver, W.G., Younghusband, H.B., and Wrigley, N.G., 1971, Purification and properties of chick embryo lethal orphan virus (CELO), *Virology* **45**:598.

Lewis, J.B., and Mathews, M.B., 1980, Control of adenovirus early gene expression: A class of "immediate early" products, *Cell* **21**:303.

Lewis, J.B., Atkins, J.F., Anderson, C.W., Baum, P.R., and Gesteland, R.F., 1975, Mapping of late adenovirus genes by cell-free translation of RNA selected by hybridization to specific DNA fragments, *Proc. Natl. Acad. Sci. U.S.A.* **72**:1344.

Lewis, J.B., Atkins, J.F., Baum, P.R., Solem, R., Gesteland, R.B., and Anderson, C.W., 1976, Location and identification of the genes for adenovirus type 2 early polypeptides, *Cell* **7**:141.

Lewis, J.B., Anderson, C.W., and Atkins, J.F., 1977, Further mapping of late adenovirus genes by cell-free translation of RNA selected by hybridization to specific DNA fragments, *Cell* **12**:37.

Lewis, J.B., Esche, H., Smart, J.E., Stillman, B.W., Harter, M.L., and Mathews, M.B., 1979, Organization and expression of the left third of the genome of adenovirus, *Cold Spring Harbor Symp. Quant. Biol.* **44**:493.

Lichy, J.H., Field, J., Horwitz, M.S., and Hurwitz, J., 1982, Separation of the adenovirus terminal protein precursor from its associated DNA polymerase: Role of both proteins in the initiation of adenovirus DNA replication, *Proc. Natl. Acad. Sci. U.S.A.* **79**:5225.

Lupker, J.H., Davis, A., Jochemsen, H., and van der Eb, A.J., 1980, *In vitro* synthesis of adenovirus type 5 T antigens. I. Translation of early region 1-specific RNA from lytically infected cells, *J. Virol.* **37**:524.

Maat, J., and van Ormondt, H., 1979, The nucleotide sequence of the transforming HindIII-G fragment of adenovirus type 5 DNA, *Gene* **6**:75.

Maat, J., van Beveren, C.P., and van Ormondt, H., 1980, The nucleotide sequence of adenovirus type 5 early region E-1: The region between map positions 8.0 (HindIII site) and 11.8 (SmaI site), *Gene* **10**:27.

Maizel, J.V., White, D.O., and Scarff, M.D., 1968, The polypeptides of adenovirus. II. Soluble proteins, cores, top components and structure of the virion, *Virology* **36**:126.

Mathews, M.B., 1980, Binding of adenovirus VA RNA to mRNA: A possible role in splicing?, *Nature (London)* **285**:575.

Mathews, M.B., and Pettersson, U., 1978, The low molecular weight RNA's of adenovirus 2 infected cells, *J. Mol. Biol.* **119**:293.

Mathis, D.J., Elkaim, R., Kedinger, C., Sassone-Corsi, P., and Chambon, P., 1981, Specific *in vitro* initiation of transcription on the adenovirus type 2 early and late EII transcription units, *Proc. Natl. Acad. Sci. U.S.A.* **78**:7383.

Mautner, V., Williams, J., Sambrook, J., Sharp, P.A., and Grodzicker, T., 1975, The location of the genes coding for hexon and fibre proteins in adenovirus DNA, *Cell* **5**:93.

Mayer, A.J., and Ginsberg, H.S., 1977, Persistence of type 5 adenovirus DNA in cells transformed by the temperature-sensitive mutant H5ts125, *Proc. Natl. Acad. Sci. U.S.A.* 74:785.

McAllister, R.M., Nicolson, M.O., Lewis, A.M., McPherson, I., and Huebner, R.J., 1969, Transformation of rat embryo cells by adenovirus type 1, *J. Gen. Virol.* 4:29.

McBride, W.D., and Weiner, A., 1964, *In vitro* transformation of hamster kidney cells by human adenovirus type 12, *Proc. Soc. Exp. Biol. Med.* 115:870.

McGrogan, M., and Raskas, H.J., 1978, Two regions of the adenovirus 2 genome specify families of late polysomal RNA's containing common sequences, *Proc. Natl. Acad. Sci. U.S.A.* 75:625.

Miller, S.J., Ricciardi, R.P., Roberts, B.E., Paterson, B.M., and Mathews, M.B., 1980, Arrangement of messenger RNAs and protein coding sequences in the major late transcription unit of adenovirus 2, *J. Mol. Biol.* 142:455.

Mirza, M.A., and Weber, J., 1982, Structure of adenovirus chromatin, *Biochim. Biophys. Acta* 696:76.

Montell, C., Fisher, E.F., Caruthers, M.H., and Berk, A.J., 1982, Resolving the functions of overlapping viral genes by site-specific mutagenesis at a mRNA splice site, *Nature (London)* 295:380.

Murray, V., and Holliday, R., 1979, Mechanism of RNA splicing for gene transcripts, *FEBS Lett.* 106:5.

Naroditsky, B.S., Kolinina, T.I., Goldberg, E.Z., Borovik, A.S., Karamov, E.V., and Tikchonenko, T.I., 1980, Analysis of DNA from human adenovirus type 6 with restriction endonucleases HindIII, BglII, and BamHI, *Biochim. Biopys. Acta* 606:214.

Nevins, J.R., 1981, Mechanism of activation of early viral transcription by the adenovirus E1A gene products, *Cell* 26:213.

Nevins, J.R., and Darnell, J.E., 1978, Groups of adenovirus type 2 mRNA's derived from a large primary transcript: Probably nuclear origin and possible common 3'-ends, *J. Virol.* 25:811.

Nevins, J.R., and Jensen-Winkler, J.J., 1980, Regulation of early adenovirus transcription: A protein product of early region 2 specifically represses region 4 transcription, *Proc. Natl. Acad. Sci. U.S.A.* 77:1893.

Nevins, J.R., Blanchard, J.M., and Darnell, J.E., 1980, Transcription units of adenovirus type 2: Termination of transcription beyond the poly[A] addition site in early regions 2 and 4, *J. Mol. Biol.* 144:377.

Nicolson, M.O., and McAllister, R.M., 1972, Infectivity of human adenovirus 1 DNA, *Virology* 48:14.

Ohe, K., 1972, Virus-coded origin of a low molecular weight RNA from KB cells infected with adenovirus 2, *Virology* 47:726.

Ohe, K., and Weissmann, S., 1970, Nucleotide sequence of an RNA from cells infected with adenovirus 2, *Science* 167:879.

Ohen, K., and Weissman, S., 1971, The nucleotide sequence of a low molecular weight RNA from cells infected with adenovirus 2, *J. Biol. Chem.* 246:6999.

Padmanabhan, R., and Green, M., 1976, Evidence for palindromic sequences near the termini of adenovirus 2 DNA, *Biochem. Biophys. Res. Commun.* 69:860.

Padmanabhan, R., and Padmanabhan, R.V., 1977, Specific interaction of a protein(s) near the termini of adenovirus 2 DNA, *Biochem. Biophys. Res. Commun.* 80:955.

Pan, J., Celma, M.L., and Weissman, S.M., 1977, Studies of low molecular weight RNA from cells infected with adenovirus 2. III. The sequence of the promoter for VA-RNA I. *J. Biol. Chem.* 252:9047.

Patch, C.T., Lewis, A.M., and Levine, A.S., 1972, Evidence for a transcriptional control region of SV40 in the adenovirus 2–SV40 hybrid, Ad2+ND1, *Proc. Natl. Acad. Sci. U.S.A.* 69:3375.

Pereira, M.S., Pereira, H.G., and Clarke, S.K., 1965, Human adenovirus type 31: A new serotype with oncogenic properties, *Lancet* 1:21.

Perricaudet, M., Akusjärvi, G., Virtanen, A., and Pettersson, U., 1979, Structure of two

spliced mRNA's from the transforming region of human subgroup C adenoviruses, *Nature (London)* 281:694.
Perricaudet, M., le Moullec, J.M., and Pettersson, U., 1980, The predicted structure of two adenovirus T-antigens, *Proc. Natl. Acad. Sci. U.S.A.* 77:3778.
Persson, H., and Philipson, L., 1982, Regulation of adenovirus gene expression, *Curr. Top. Microbiol. Immunol.* 97:157.
Persson, H., Pettersson, U., and Mathews, M.B., 1978, Synthesis of a structural polypeptide in the absence of viral DNA replication, *Virology* 90:67.
Persson, H., Mathisen, B., Philipson, L., and Pettersson, U., 1979a, A maturation protein in adenovirus morphogenesis, *Virology* 93:198.
Persson, H., Signas, C., and Philipson, L., 1979b, Purification and characterization of an early glycoprotein from adenovirus type 2 infected cells, *J. Virol.* 29:938.
Persson, H., Jansson, M., and Philipson, L., 1980a, Synthesis and genomic site for an adenovirus type 2 early protein, *J. Mol. Biol.* 136:375.
Persson, H., Jörnvall, H., and Zabielski, J., 1980b, Multiple mRNA species for the precursor to an adenovirus-encoded glycoprotein: Identification and structure of the signal sequence, *Proc. Natl. Acad. Sci. U.S.A.* 77:6349.
Pettersson, U., and Mathews, M.B., 1977, The gene and the messenger RNA for adenovirus polypeptide IX, *Cell* 12:741.
Piña, M., and Green, M., 1965, Biochemical studies on adenovirus multiplication. IX. Chemical and base composition analysis of 28 human adenoviruses, *Proc. Natl. Acad. Sci. U.S.A.* 54:547.
Ponomareva, T.I., Grodnitskaya, N.A., Goldberg, E.E., Chaplygina, N.M., Naroditsky, B.S., and Tikchonenko, T.I., 1979, Biological activity of intact and cleaved DNA of the simian adenovirus 7, *Nucleic Acids Res.* 5:3119.
Prage, L., and Pettersson, U., 1971, Structural proteins of adenoviruses. VII. Purification and properties of an arginine-rich core protein from adenovirus type 2 and type 3, *Virology* 45:364.
Prage, L., Pettersson, U., and Philipson, L., 1968, Internal base proteins in adenovirus, *Virology* 36:508.
Prage, L., Pettersson, U., Hoglund, S., Lonberg-Holm, K., and Philipson, L., 1970, Structural proteins of adenoviruses. IV. Sequential degradation of the adenovirus type 2 virion, *Virology* 42:341.
Price, R., and Penman, S., 1972, A distinct RNA polymerase activity, synthesizing 5.5S, 5S and 4S RNA in nuclei from adenovirus 2 infected HeLa cells, *J. Mol. Biol.* 70:435.
Rekosh, D., 1981, Analysis of the DNA–terminal protein from different serotypes of human adenovirus, *J. Virol.* 40:329.
Rekosh, D.M.K., Russell, W.C., Bellett, A.J.D., and Robinson, A.J., 1977, Identification of a protein linked to the ends of adenovirus DNA, *Cell* 11:283.
Rijnders, A.W.M., van Maarschalkerweerd, M.W., Visser, L., Reemst, A.M.C.B., Sussenbach, J.S., and Rozijn, T.H., 1983, Expression of integrated viral DNA sequences outside the transforming region in eight adenovirus-transformed cell lines, *Biochim. Biophys. Acta* 739:48.
Roberts, R.J., 1981, Restriction and modification enzymes and their recognition sequences, *Nucleic Acids Res.* 9:r75.
Roberts, R.J., Arrand, J.R., and Keller, W., 1974, The length of the terminal repetition in adenovirus-2 DNA, *Proc. Natl. Acad. Sci. U.S.A.* 71:3829.
Robinson, A.J., and Bellett, A.J.D., 1975a, A circular DNA–protein complex from adenovirus and its possible role in DNA replication, *Cold Spring Harbor Symp. Quant. Biol.* 39:523.
Robinson, A.J., and Bellett, A.J.D., 1975b, Complementary strands of CELO virus DNA, *J. Virol.* 15:458.
Robinson, A.J., Younghusband, H.B., and Bellett, A.J.D., 1973, A circular DNA–protein complex from adenoviruses, *Virology* 56:54.
Roninson, I., and Padmanabhan, R., 1980, Studies on the nature of the linkage between the terminal protein and the adenovirus DNA, *Biochem. Biophys. Res. Commun.* 94:398.

Rósen, L., 1960, A hemagglutination-inhibition technique for typing adenoviruses, *Am. J. Hyg.* **71**:120.

Ross, S., Flint, S.J., and Levine, A.J., 1980, Identification of the adenovirus early proteins and their genomic positions, *Virology* **100**:419.

Russell, W.C., McIntosh, K., and Skehel, J.J., 1971, The preparation and properties of adenovirus cores, *J. Gen. Virol.* **11**:35.

Sambrook, J., Williams, J.F., Sharp, P.A., and Grodzicker, T., 1975, Physical mapping of temperature-sensitive mutations of adenoviruses. *J. Mol. Biol.* **97**:369.

Schwarz, E., Reinke, C., Yamamoto, N., and Zur Hausen, H., 1982, Terminal rearrangements in the genome of adenovirus type 12 mutants adapted to growth in two human tumor cell lines, *Virology* **116**:284.

Seghal, P., Fraser, N., and Darnell, J., 1979, Early Ad2 transcription units: Only promoter-proximal RNA continues to be made in the presence of DRB, *Virology* **94**:185.

Sekikawa, K., Shiroki, K., Shimojo, H., Ojima, S., and Fujinaga, K., 1978, Transformation of a rat cell line by an adenovirus 7 DNA fragment *Virology* **88**:1.

Sharp, P.A., Gallimore, P.H., and Flint, S.J., 1975, Mapping of adenovirus 2 RNA sequences in lytically infected cells and transformed cell lines, *Cold Spring Harbor Symp. Quant. Biol.* **39**:457.

Sharp, P.A., Moore, C., and Haverty, J., 1976, The infectivity of adenovirus 5–DNA protein complex, *Virology* **75**:442.

Shaw, A.R., and Ziff, E.B., 1980, Transcripts from the adenovirus-2 late promoter yields a single family of coterminal mRNA's during early infection and five families at late times, *Cell* **22**:905.

Shinagawa, M., and Padmanabhan, R., 1979, Nucleotide sequences at the inverted terminal repetition of Ad2 DNA, *Biochem. Biophys. Res. Commun.* **87**:671.

Shinagawa, M., and Padmanabhan, R., 1980, Comparative sequence analysis of the inverted terminal repetitions from different adenoviruses, *Proc. Natl. Acad. Sci. U.S.A.* **77**:3831.

Shinagawa, M., Padmanabhan, R.V., and Padmanabhan, R., 1980, The nucleotide sequence of the right-hand terminal SmaI-K fragment of adenovirus type 2 DNA, *Gene* **9**:99.

Smart, J.E., and Stillman, B.W., 1982, Adenovirus terminal protein precursor: Partial amino acid sequence and the site of covalent linkage to virus DNA, *J. Biol. Chem.* **257**:13499.

Söderland, H., Pettersson, U., Vennström, B., Philipson, L., and Mathews, M.B., 1976, A new species of virus-coded low molecular weight RNA from cells infected with adenovirus type 2, *Cell* **7**:585.

Spector, D.J., McGrogan, M., and Raskas, H.J., 1978, Regulation of the appearance of cytoplasmic RNA's from region 1 of the adenovirus genome, *J. Mol. Biol.* **126**:395.

Spector, D.J., Crossland, L.D., Halbert, D.N., and Raskas, H.J., 1980a, A 28 K polypeptide is the translation product of 9 S RNA encoded by region 1A of adenovirus 2, *Virology* **102**:218.

Spector, D.J., Halbert, D.N., Crossland, L.D., and Raskas, H.J., 1980b, Expression of genes from the transforming region of adenovirus, *Cold Spring Harbor Symp. Quant. Biol.* **44**:437.

Steenbergh, P.H., and Sussenbach, J.S., 1979, The nucleotide sequence of the right-hand terminus of adenovirus type 5 DNA: Implications for the mechanism of DNA replication, *Gene* **6**:307.

Steenbergh, P.H., Maat, J., van Ormondt, H., and Sussenbach, J.S., 1977, The nucleotide sequence at the termini of adenovirus type 5 DNA, *Nucleic Acids Res.* **4**:4371.

Stenlund, A., Perricaudet, M., Tiollais, P., and Pettersson, U., 1980, Construction of restriction enzyme fragment libraries containing DNA from human adenovirus types 2 and 5, *Gene* **10**:47.

Stillman, B.W., Lewis, J.B., Chow, L.T., Mathews, M.B., and Smart, J.E., 1981, Identification of the gene and mRNA for the adenovirus terminal protein precursor, *Cell* **23**:497.

Sugisaka, H., Sugimoto, K., Takanami, M., Shiroki, K., Saito, I., Shimojo, H., Sawada, Y., Uemizu, Y., Uesigi, S.-I., and Fujinaga, K., 1980, Structure and gene organization in the transforming HindIII-G fragment of Ad12, *Cell* **20**:777.

Sussenbach, J.S., Ellens, D.J., and Jansz, H.S., 1973, Studies on the mechanism of adenovirus

DNA replication. II. The nature of single-stranded DNA in replicative intermediates, *J. Virol.* **12**:1131.
Tamanoi, F., and Stillman, B.W., 1982, Function of adenovirus terminal protein in the initiation of DNA replication, *Proc. Natl. Acad. Sci. U.S.A.* **79**:2221.
Tate, V.E., and Philipson, L., 1979, Parental adenovirus DNA accumulates in nucleosome-like structures in infected cells, *Nucleic Acids Res.* **6**:2769.
Temple, M., Antoine, G., Delius, H., Stahl, S., and Winnacker, E.L., 1981, Replication of mouse adenovirus strain FL DNA, *Virology* **109**:1.
Tibbetts, C., 1977, Physical organization of subgroup B human adenoviruses genomes, *J. Virol.* **24**:564.
Tibbetts, C., and Pettersson, U., 1974, Complementary strand-specific sequences from unique fragments of adenovirus type 2 DNA for hybridization mapping experiments, *J. Mol. Biol.* **88**:767.
Tibbetts, C., Johansson, K., and Philipson, L., 1973, Hydroxylapatite chromatography and formamide denaturation of adenovirus DNA, *J. Virol.* **12**:218.
Tibbetts, C., Pettersson, U., Johansson, K., and Philipson, L., 1974, Transcription of the adenovirus type 2 genome. I. Relationship of cytoplasmic RNA to the complementary strands of viral DNA, *J. Virol.* **13**:370.
Tokunaga, O., Shinagawa, M., and Padmanabhan, R., 1982, Physical mapping and sequence analysis at the inverted terminal repetition of adenovirus type 4, *Gene* **18**:329.
Tolun, A., Alestrőm, P., and Pettersson, U., 1979, Sequence of inverted terminal repetitions from different adenoviruses: Demonstration of conserved sequences and homology between SA7 and SV40 DNA, *Cell* **17**:705.
Tooze, J. (ed.), 1981, *The Molecular Biology of Tumor Viruses*, 2nd rev. ed., Part 10B, *DNA Tumor Viruses*, Cold Spring Harbor Press, Cold Spring Harbor, New York.
Trentin, J.J., Yabe, Y., and Taylor, G., 1962, The quest for human cancer viruses, *Science* **137**:835.
van Beveren, C.P., Maat, J., Dekker, B.M.M., and van Ormondt, H., 1981, The nucleotide sequence of the gene for protein IVa2 and of the 5′ leader segment of the major late mRNA's of adenovirus type 5, *Gene* **16**:179.
van den Elsen, P., de Pater, S., Houweling, A., Van der Veer, J., and van der Eb, A.J., 1982, The relationship between region E1A and E1B of human adenoviruses in cell transformation, *Gene* **18**:175.
van der Eb, A.J., and van Kesteren, L.W., 1966, Structure and molecular weight of the DNA of adenovirus type 5, *Biochim. Biopys. Acta* **129**:441.
van der Eb, A.J., van Kesteren, L.W., and van Bruggen, E.F.J., 1969, Structural properties of adenovirus DNAs, *Biochim. Biopys. Acta* **182**:530.
van der Eb, A.J., Mulder, C., Graham, F., and Houweling, A., 1977, Transformation with specific fragments of adenovirus DNA. I. Isolation of specific fragments with transforming activity of adenoviruses 2 and 5 DNA, *Gene* **2**:115.
van der Eb, A.J., van Ormondt, H., Schrier, P.I., Lupker, J.H., Jochemsen, H., van den Elsen, P.J., Deleys, R.J., Maat, J., van Beveren, C.P., Dijkema, R., and de Waard, A., 1979, Structure and function of the transforming genes of human adenoviruses and SV40, *Cold Spring Harbor Symp. Quant. Biol.* **44**:383.
van der Vliet, P.C., and Levine, A.J., 1973, DNA-binding proteins specific for cells infected by adenovirus, *Nature* (*London*) *New Biol.* **246**:170.
van der Vliet, P.C., and Sussenbach, J.S., 1975, An adenovirus type 5 gene function required for initiation of viral DNA replication, *Virology* **67**:415.
van der Vliet, P.C., Levine, A.J., Ensinger, M., and Ginsberg, H.S., 1975, Thermolabile DNA-binding proteins from cells infected with a temperature-sensitive mutant of adenovirus defective in viral DNA synthesis, *J. Virol.* **15**:348.
van der Vliet, P.C., Zandberg, J., and Jansz, H.S., 1977, Evidence for a function of the adenovirus DNA-binding protein in initiation of DNA synthesis as well as in elongation of nascent DNA chains, *Virology* **80**:98.
van Ormondt, H., Maat, J., de Waard, A., and van der Eb, A.J., 1978, The nucleotide sequence of the transforming HpaI-E fragment of adenovirus type 5 DNA, *Gene* **4**:309.

van Ormondt, H., Maat, J., and van Beveren, C.P., 1980a, The nucleotide sequence of the transforming region E-1 of adenovirus type 5 DNA, *Gene* **11**:299.

van Ormondt, H., Maat, J., and Dijkema, R., 1980b, Comparison of nucleotide sequences of the early E1A regions for subgroups A, B and C of human adenoviruses, *Gene* **12**:63.

van Wielink, P.S., 1978, A DNA–protein complex from adenovirus type 5 and its possible role in viral DNA replication, Academic thesis, State University of Utrecht, Utrecht.

Varsanyi, T.M., Winberg, G., and Wadell, G., 1977, DNA restriction site mapping of adenovirus type 16 with BamI, EcoRI, HpaI and SalI. *FEBS Lett.* **76**:151.

Vennström, B., Pettersson, U., and Philipson, L., 1978a, Initiation of transcription in nuclei isolated from adenovirus-infected cells, *Nucleic Acids Res.* **5**:205.

Vennström, B., Pettersson, U., and Philipson, L., 1978b, Two initiation sites for adenovirus 5.5S RNA, *Nucleic Acids Res.* **5**:195.

Virtanen, A., Pettersson, U., Le Moullec, J.M., Tiollais, P., and Perricaudet, M., 1982a, Different mRNAs from the transforming region of highly oncogenic and non-oncogenic human adenoviruses, *Nature* [*London*] **295**:705.

Virtanen, A., Alestrőm, P., Persson, H., Katze, M.G., and Pettersson, U., 1982b, An adenovirus agnogene, *Nucleic Acids Res.* **10**:2539.

Vlak, J.M., Rozijn, T.H., and Sussenbach, J.S., 1975, Studies on the mechanism of replication of adenovirus DNA. IV. Discontinuous DNA chain propagation, *Virology* **63**:168.

Wadell, G., 1978, Classification of human adenoviruses by SDS–polyacrylamide gel electrophoresis of structural polypeptides, *Intervirology* **11**:47.

Weber, J., Bégin, M., and Khitto, G., 1975, Genetic analysis of adenovirus type 2. II. Preliminary phenotypic characterization of temperature-sensitive mutants, *J. Virol.* **15**:1049.

Weinmann, R., Raskas, H.J., and Roeder, R.G., 1974, Role of DNA-dependent RNA polymerases II and III in transcription of the adenovirus genome late in productive infection, *Proc. Natl. Acad. Sci. U.S.A.* **71**:3436.

Weinmann, R., Brendler, T.G., Raskas, H.J., and Roeder, R.G., 1976, Low molecular weight viral RNA's transcribed by RNA polymerase III during adenovirus 2 infection, *Cell* **7**:5577.

Westphal, H., and Lai, S.-P., 1977, Quantitative electron microscopy of adenovirus RNA, *J. Mol. Biol.* **116**:525.

Westphal, H., Meyer, J., and Maizel, J.V., Jr., 1976, Mapping of adenovirus 2 mRNA by electron microscopy, *Proc. Natl. Acad. Sci. U.S.A.* **73**:2069.

Wilson, M.C., Fraser, N.W., and Darnell, J.E., 1979, Mapping of DNA initiation sites by high doses of ultraviolet irradiation: Evidence for three independent promoters within the left 11% of the Ad-2 genome, *Virology* **94**:15.

Winberg, G., and Hammarskjöld, M.L., 1980, Isolation of DNA from agarose gels using DEAE paper: Application to restriction site mapping of adenovirus type 16 DNA, *Nucleic Acids Res.* **8**:253.

Wolfson, J., and Dressler, D., 1972, Adenovirus DNA contains an inverted terminal redundancy, *Proc. Natl. Acad. Sci. U.S.A.* **69**:3054.

Wu, M., Roberts, R.J., and Davidson, N., 1977, Structure of the inverted terminal repetition of adenovirus type 2 DNA, *J. Virol.* **21**:766.

Younghusband, H.B., and Bellett, A.J.D., 1971, Mature form of DNA from chicken embryo lethal orphan virus, *J. Virol.* **8**:265.

Zain, B.S., and Roberts, R.J., 1979, Sequences from the beginning of the fiber messenger RNA of adenovirus-2, *J. Mol. Biol.* **131**:341.

Zain, S., Sambrook, J., Roberts, R.J., Keller, W., Fried, M., and Dunn, A.R., 1979a, Nucleotide sequence analysis of the leader segments in a cloned copy of adenovirus 2 fiber mRNA, *Cell* **16**:851.

Zain, S., Gingeras, T.R., Bullock, P., Wong, G., and Gelinas, R.E., 1979b, Determination and analysis of adenovirus-2 DNA sequences which may include signals for late mRNA processing, *J. Mol. Biol.* **135**:413.

Ziff, E.B., and Evans, R.M., 1978, Coincidence of the promoter and capped 5′ terminus of RNA from the adenovirus 2 major late transcription unit, *Cell* **155**:1463.

Ziff, E.B., and Fraser, N., 1978, Adenovirus type 2 late mRNA; Structural evidence for 3′-coterminal species. *J. Virol.* **25**:897.